



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Redes neuronales recurrentes para la predicción de casos por Covid-19 usando datos abiertos en el Perú, 2020

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniero de Sistemas

AUTORES:

Catunta Esquivel, Saulo (ORCID: 0000-0003-3408-3601)

Roncal Hidalgo, Dalyuska Mabel (ORCID: 0000-0002-6692-6530)

ASESOR:

Mg. Milner David Liendo Arévalo (ORCID: 0000-0002-7665-361X)

LÍNEA DE INVESTIGACIÓN:

Sistema de Información y Comunicaciones

LIMA – PERÚ

2020

Dedicatoria

Dedicamos esta tesis a nuestros queridos padres quienes nos han apoyado siempre, dándonos aliento y motivación para no rendirnos y a Dios que guía nuestro camino hacia el éxito.

Agradecimiento

Agradecemos encarecidamente a nuestros docentes de la facultad de Ingeniería de la Universidad César Vallejo, en especial al Mg. Milner David Liendo Arévalo que nos apoyó en la culminación de esta investigación con su asesoramiento y orientación profesional.

Índice de contenidos

Carátula.....	i
Dedicatoria.....	ii
Agradecimiento.....	iii
Índice de contenidos.....	iv
Índice de tablas.....	v
Índice de figuras.....	vii
Resumen.....	ix
Abstract.....	x
I. INTRODUCCIÓN	1
II. MARCO TEÓRICO.....	6
III. MÉTODO	23
3.1 Tipo y diseño de investigación.....	24
3.2 Variables y operacionalización.....	25
3.3 Población, muestra y muestreo	25
3.4 Técnicas e instrumentos de recolección de datos	27
3.5 Procedimientos.....	28
3.5 Método de análisis de datos.....	28
3.7 Aspectos éticos	29
IV. RESULTADOS	30
4.1 Contraste de hipótesis.....	31
V. DISCUSIÓN	39
VI. CONCLUSIONES.....	42
VII. RECOMENDACIONES	45
REFERENCIAS	47
ANEXOS	55

Índice de tablas

TABLA 1 CUADRO COMPARATIVO METODOLOGÍAS DE MINERÍA DE DATOS	21
TABLA 2: ESTADÍSTICOS DESCRIPTIVOS DE LA TASA DE LETALIDAD DE LOS DATOS Y DE LA PREDICCIÓN DE COVID-19.....	31
TABLA 3: PRUEBA DE NORMALIDAD DE LA TASA DE MORTALIDAD DE LOS DATOS Y DE LA PREDICCIÓN DE COVID-19.....	32
TABLA 4: PRUEBA T-STUDENT PARA LA TASA DE MORTALIDAD DE LOS DATOS Y PREDICCIÓN DE COVID-19	34
TABLA 5: ESTADÍSTICA DESCRIPTIVA DE LA TASA DE LETALIDAD DE LOS DATOS Y DE LA PREDICCIÓN DE COVID-19.....	35
TABLA 6: PRUEBA DE NORMALIDAD DE LA TASA DE LETALIDAD DE LOS DATOS Y DE LA PREDICCIÓN DE COVID-19	36
TABLA 7:PRUEBA T-STUDENT PARA LA TASA DE LETALIDAD DE LOS DATOS Y PREDICCIÓN DE COVID-19	38
TABLA 8: MATRIZ DE OPERACIONALIZACIÓN DE VARIABLES.....	56
TABLA 9 MATRIZ DE CONSISTENCIA.....	58
TABLA 10 SOFTWARE.....	63
TABLA 11 FRAMEWORK.....	64
TABLA 12 LENGUAJE.....	64
TABLA 13 LIBRERÍAS	64
TABLA 14 HOSTING.....	65
TABLA 15 ROLES Y RESPONSABLES	66
TABLA 16 HISTORIA DE USUARIO HU01	67
TABLA 17 HISTORIA DE USUARIO HU02	67
TABLA 18 HISTORIA DE USUARIO HU03.....	68
TABLA 19 HISTORIA DE USUARIO HU04	68
TABLA 20 HISTORIA DE USUARIO HU05	69
TABLA 21 HISTORIA DE USUARIO HU06	69
TABLA 22 HISTORIA DE USUARIO HU07	70
TABLA 23 HISTORIA DE USUARIO HU08.....	70
TABLA 24 HISTORIA DE USUARIO HU09	71
TABLA 25 HISTORIA DE USUARIO HU10	71
TABLA 26 HISTORIA DE USUARIO HU11	72
TABLA 27 HISTORIA DE USUARIO HU12	72
TABLA 28 HISTORIA DE USUARIO HU13	73
TABLA 29 HISTORIA DE USUARIO HU14.....	73
TABLA 30 HISTORIA DE USUARIO HU15	74
TABLA 31 PILA DE PRODUCTO.....	74
TABLA 32 SPRINT GENERAL.....	76
TABLA 33 SPRINT 1 ENTRADA	77
TABLA 34 RESUMEN SPRINT 1	79
TABLA 35 SPRINT 1 SALIDA	81
TABLA 36 SPRINT 2 ENTRADA	81
TABLA 37 RESUMEN SPRINT 2	84
TABLA 38 SPRINT 2 SALIDA	85
TABLA 39 SPRINT 3 ENTRADA	85
TABLA 40 RESUMEN SPRINT 3	88

TABLA 41 SPRINT 3 SALIDA	90
TABLA 42 SPRINT 4 ENTRADA	90
TABLA 43 RESUMEN SPRINT 4	93
TABLA 44 SPRINT 4 SALIDA	95
TABLA 45 SELECCIÓN DE METODOLOGÍA POR JUICIO DE EXPERTOS	96
TABLA 46 FALLECIDOS COVID-19 - PREDICCIÓN.....	106
TABLA 47 COMPARACIÓN KDD - SEMMA	120
TABLA 48 COMPARACIÓN KDD - CRISP-DM	120
TABLA 49 COMPARACIÓN CRIP-DM - SEMMA.....	120

Índice de figuras

FIGURA 1: DISEÑO NO EXPERIMENTAL - TRANSVERSAL	24
FIGURA 2: PROCEDIMIENTO DE PROYECTO	28
FIGURA 3: MÉTODO DE ANÁLISIS DE DATOS	28
FIGURA 4: GRAFICO DE BARRAS DE LA TASA DE MORTALIDAD DE LOS DATOS Y PREDICCIÓN DE COVID-19	32
FIGURA 5: HISTOGRAMA DE LA TASA DE MORTALIDAD USANDO LOS DATOS DEL COVID-19	33
FIGURA 6: HISTOGRAMA DE LA TASA DE MORTALIDAD DE LA PREDICCIÓN DEL COVID-19	33
FIGURA 7: PRUEBA DE T-STUDENT DE LA TASA DE MORTALIDAD	34
FIGURA 8: GRAFICO DE BARRAS DE LA TASA DE MORTALIDAD DE LOS DATOS Y PREDICCIÓN DE COVID-19	36
FIGURA 9: HISTOGRAMA DE LA TASA DE LETALIDAD USANDO LOS DATOS DEL COVID-19	37
FIGURA 10: HISTOGRAMA DE LA TASA DE LETALIDAD USANDO LOS DATOS DEL COVID-19	37
FIGURA 11: PRUEBA DE T-STUDENT DE LA TASA DE LETALIDAD	38
FIGURA 12 REGISTRO TASA DE MORTALIDAD DE LOS DATOS DE COVID-19	59
FIGURA 13 TASA DE LETALIDAD DE LOS DATOS DE COVID 19	60
FIGURA 14 TASA DE MORTALIDAD DE PREDICCIÓN DE COVID 19	61
FIGURA 15 TASA DE LETALIDAD DE PREDICCIÓN DE COVID 19	62
FIGURA 16 PREDICCIÓN DE PERSONAS CONFIRMADAS CON COVID 19	77
FIGURA 17 PREDICCIÓN DE PERSONAS FALLECIDAS POR COVID 19	78
FIGURA 18 ACTUALIZAR DATOS DE LOS CASOS CONFIRMADOS DE COVID 19	78
FIGURA 19 ACTUALIZAR DATOS DE LOS CASOS FALLECIDOS POR COVID 19	79
FIGURA 20 BURN DOWN CHART ESTIMADO SPRINT 1	80
FIGURA 21 BURN DOWN CHART REAL SPRINT 1	80
FIGURA 22 GENERAR GRÁFICO DE SERIE DE TIEMPO CON LOS REGISTROS DE CASOS CONFIRMADOS POR DÍA CON COVID 19	82
FIGURA 23 GENERAR GRÁFICO DE SERIE DE TIEMPO CON LOS REGISTROS DE CASOS CONFIRMADOS ACUMULADOS CON COVID 19	82
FIGURA 24 GENERAR GRÁFICO DE SERIE DE TIEMPO CON LOS REGISTROS DE CASOS FALLECIDOS POR DÍA POR COVID 19	83
FIGURA 25 GENERAR GRÁFICO DE SERIE DE TIEMPO CON LOS REGISTROS DE CASOS FALLECIDOS ACUMULADOS POR COVID 19	83
FIGURA 26 BURN DOWN CHART ESTIMADO SPRINT 2	84
FIGURA 27 BURN DOWN CHART REAL SPRINT 2	84
FIGURA 28 MOSTRAR TABLA DE LOS CASOS CONFIRMADOS CON COVID-19	86
FIGURA 29 MOSTRAR TABLA DE LOS CASOS FALLECIDOS POR COVID-19	87
FIGURA 30 MOSTRAR TABLA DE TASA DE MORTALIDAD DE COVID-19 POR CADA 10 000 HABITANTES	87
FIGURA 31 MOSTRAR TABLA DE LA TASA DE LETALIDAD DE COVID-19	88
FIGURA 32 BURN DOWN ESTIMADO SPRINT 3	89
FIGURA 33 BURN DOWN CHART REAL SPRINT 3	89
FIGURA 34 SELECCIÓN DE DEPARTAMENTO O PERÚ PARA PREDECIR	92
FIGURA 35 MOSTRAR LOS CASOS TOTALES DE COVID--19 DEL MINSA	92
FIGURA 36 MOSTRAR GRÁFICO CIRCULAR DE LOS CASOS DE COVID-19	93
FIGURA 37 BURN DOWN CHART ESTIMADO SPRINT 4	94
FIGURA 38 BURN DOWN CHART REAL SPRINT 4	94
FIGURA 39 CASOS CONFIRMADOS	98
FIGURA 40 CASOS FALLECIDOS	99
FIGURA 41 DESCRIPCIÓN DATOS CONFIRMADOS	99

FIGURA 42 DETALLE DATOS CONFIRMADOS	99
FIGURA 43 DESCRIPCIÓN DE LOS DATOS DE FALLECIDOS	100
FIGURA 44 DETALLE DATOS DE FALLECIDOS.....	100
FIGURA 45 CONFIRMADOS POR COVID-19.....	101
FIGURA 46 DATOS DE CONFIRMADOS POR COVID-19 SELECCIÓN DE COLUMNAS DE DATOS CONFIRMADOS.....	102
FIGURA 47 CANTIDAD DE CASOS PARA CADA FECHA Y DEPARTAMENTO	103
FIGURA 48 CASOS DE COVID-19 DEL PERÚ	103
FIGURA 49 CANTIDAD DE CASOS FALLECIDOS POR COVID-19.....	104
FIGURA 50 CASOS DE COVID-19 DEL PERÚ- FALLECIDOS	104
FIGURA 51 DISEÑO DE PRUEBA	105
FIGURA 52 ENTRENAMIENTO DE LA RED NEURONAL RECURRENTE PARA FALLECIDOS	106
FIGURA 53 RESULTADO RMSE PARA FALLECIDOS	107
FIGURA 54 CONFIRMADOS CON COVID-19 - PREDICCIÓN	107
FIGURA 55 ENTRENAMIENTO DE LA RED NEURONAL RECURRENTE PARA CONFIRMADOS.....	107
FIGURA 56 RESULTADOS RMSE CONFIRMADOS.....	108
<i>FIGURA 57 PREFERENCIA DE METODOLOGÍA</i>	<i>121</i>
FIGURA 58 RESULTADO DE TURNITIN	122

Resumen

El problema de la investigación fue cómo utilizar redes neuronales recurrentes para la predicción de casos por Covid-19 usando datos abiertos en el Perú, 2020. El objetivo de la investigación fue utilizar las redes neuronales recurrentes para predecir los casos por Covid-19 usando datos abiertos en el Perú, 2020. La metodología de investigación es de tipo aplicada, con un enfoque cuantitativo y diseño no experimental de corte transversal. Para el desarrollo del proyecto se utilizó la metodología SCRUM. En resultados se obtuvo un aumento medio de la tasa de mortalidad de Covid-19 por cada 10 000 habitantes entre los datos recogidos (10.9818) y la predicción (11,1057), lo cual muestra que el aumento de la tasa de mortalidad ha sido de 0,1248 por cada 10 000 habitantes. Y en la variación de la tasa de letalidad de Covid-19 entre los datos recogidos (3,7335) y la predicción (3,7036), lo cual indica que hubo una variación media del 0.03 % en la tasa de letalidad. Se concluye que al utilizar las redes neuronales recurrentes en específico las RNN híbridas permiten obtener una predicción de casos confirmados como de fallecidos que a pesar de tener una cantidad limitada de entradas de registros para su entrenamiento permite proporcionar cifras que se acercan a la realidad para que pueda sacar la tasa de letalidad y mortalidad.

Palabras clave: RNN, modelo predictivo, Covid-19, BI-LSTM, BI-GRU

Abstract

The research problem was how to use recurrent neural networks to predict Covid-19 cases using open data in Peru, 2020. The objective of the research was to use recurrent neural networks to predict Covid-19 cases using open data in Peru, 2020. The research methodology is applied, with a quantitative approach and a non-experimental cross-sectional design. For the development of the project, the SCRUM methodology was used. The results obtained an average increase in the mortality rate of Covid-19 per 10 000 inhabitants between the data collected (10.9818) and the prediction (11,1057), which shows that the increase in the mortality rate It has been 0.1248 per 10,000 inhabitants. And in the variation of the fatality rate of Covid-19 between the data collected (3.7335) and the prediction (3.7036), which indicates that there was an average variation of 0.03% in the fatality rate. It is concluded that by using the recurrent neural networks specifically, the hybrid NRNs allow obtaining a prediction of confirmed cases and deaths that, despite having a limited number of record entries for their training, allows providing figures that are close to reality so that can get the fatality and mortality rate.

Keywords: RNN, predictive model, Covid-19, BI-LSTM, BI-GRU

I. INTRODUCCIÓN

En la actualidad la pandemia mundial causada por el Covid-19 ha cambiado la forma de vida de las personas en todo el mundo. El impacto que esto ha generado es de gran magnitud, considerando que no es la primera pandemia que se experimenta en el mundo, en la historia el virus más letal es conocidos como Variola virus (viruela), causando aproximadamente 300 millones de humanos muertos en el mundo.

La pandemia que se ha estado presentando en los últimos meses ha afectado al mundo, “el nuevo coronavirus 2019 (2019-nCoV / SARS-CoV-2) ha dado lugar a un brote de neumonía viral en Wuhan, China desde diciembre de 2019. La Organización Mundial de la Salud (OMS) ahora ha denominado la enfermedad Coronavirus Disease 2019 (COVID-19)” (Wang *et al.*, 2020). Por otra parte, indica que “hasta el 13 de abril de 2020, se han confirmado más de 1,800,000 casos, especialmente en los Estados Unidos (558,526 casos), Italia (156,363 casos) y España (169,496 casos), presentando una tendencia tenue en la prevención y control de enfermedades efectivas” (Zhao *et al.*, 2020).

Por otro lado, se indicó que se reportaron varios casos de neumonía en la ciudad Wuhan, la cual se encuentra localizada en China. Según las autoridades de sanidad de China indican que la fuente del virus fueron los animales vendidos en el mercado de Huanan (CNN Español, 2020).

El 24 de abril del 2020, se obtuvo que más de 2 millones 761 mil 121 personas a nivel mundial dieron positivo en las pruebas de Covid-19, de los cuales el 65.5% siguen infectados, el 27.5% se recuperaron y el 7% fallecieron. Las consecuencias del coronavirus fueron fatales, ya que se expandió mundialmente abarcando muchos casos de contagio y muertes (CSSE, 2020).

Hasta el 23 de febrero del año 2020 se ha comprobado que la tasa de mortalidad es del 2% al 4% en Wuhan, y del 0,7% fuera de Wuhan (OMS, 2020).

En Estados Unidos se confirmó el primer caso de Covid-19 en los primeros días de enero del 2020, su actual presidente Donald Trump todavía no tomaba las medidas necesarias para controlar el virus, tiempo después el país primera potencia mundial se convirtió en el nuevo epicentro mundial de la pandemia de covid-19 (BBC, 2020).

Hasta la fecha 24 de abril del 2020 EEUU tuvo 890 mil .524 casos confirmados, de los cuales el 83.4% siguen infectados, el 10.9% se recuperaron y el 5.7% fallecieron (CSSE, 2020a).

En el Perú desde el 16 de marzo de 2020 se tomó la medida de emergencia de mantener a los ciudadanos en cuarentena para poder minimizar los riesgos de contagio, pero igualmente hasta la fecha 24 de abril del 2020, se presenta una cantidad de 21648 de personas infectadas con el virus y 634 de personas fallecidas (MINSA, 2020).

El fin de la presente investigación es aportar conocimientos sobre cómo realizar predicciones mediante las redes neuronales recurrentes para los casos futuros de Covid-19 obteniendo resultados confiables del número de casos confirmados y fallecidas. Cuando se hace referencia a la relevancia que cuenta la investigación de un problema en la formación de una teoría científica. Ello llega a implicar que el estudio va a aceptar, elaborar una innovación científica por lo que es importante realizar un balance o estado de la cuestión del problema que se está investigando; manifestar si va a ser necesario para rebatir los resultados de distintas investigaciones en debate o extender un modelo teórico. (Ñaupas, Palacios, Valdivia y Romero, 2018).

El presente trabajo se justifica de manera metodológica, porque se desarrollará utilizando un enfoque cuantitativo de tipo aplicada con un diseño no experimental y transversal. Además, para la recolección de datos se utilizarán la información registrada en la Plataforma Nacional de Datos Abiertos del Perú, el cual contribuirá en la obtención de datos que serán estructurados con el fin de mostrar esquemas estadísticos con la que se puede sondear o predecir circunstancias positivas o negativas, y a partir de eso tomar las medidas necesarias para controlar la situación. Por otra parte, declararon que “al mencionar los instrumentos y técnicas de investigación que se utilizaron, puede secundar a otras investigaciones de temática similar. La cuales pueden ser técnicas o instrumentos peculiares, como cuestionarios, diagramas de muestreo, modelos, pruebas de hipótesis, test, etc.” (Ñaupas, Palacios, Valdivia y Romero, 2018a).

Esta investigación se justifica de manera práctica, porque los resultados que se obtengan gracias a las redes neuronales recurrentes permitirán aportar en muchos campos de investigación, de las cuales aportará una visión a futuro de los niveles de peligro y por ello se podrá tomar las medidas preventivas. Los datos predictivos que se podrán obtener gracias las redes neuronales recurrentes brindarán un panorama más amplio de la situación futura que se tendrá a causa del Covid-19, específicamente con los datos del número de personas contagiadas y fallecidas. Esto podría formar parte de nuevas investigaciones que apoyan con el desarrollo de nuevos modelos predictivos. “Se refiere fundamentalmente a explicar, la factibilidad, viabilidad y utilidad de los nuevos conocimientos obtenidos: así como la pertinencia social, tecnológica, económica, científica y política” (Solíz Desiderio, 2019).

El presente trabajo permitirá una visión del estado en que se encontraría la sociedad frente al Covid-19. Buscando apoyar a que las personas estén informadas y tengan conocimiento del nivel de riesgo en que se pueda encontrar la población, permitirá tener una versión a futuro de los niveles de peligro. Por otro lado, indica que este criterio de relevancia social, implica dar solución a problemas que atañen a la sociedad y su entorno, porque la situación problemática del Tema, está afectando más específicamente a ciertos grupos de la población [...]. Finalmente, el criterio de relevancia social, está determinado por la naturaleza y pertinencia del problema que le permite: explorar analizar, explicar y transformar los resultados obtenidos en la Investigación Científica (Solíz, 2019).

Sobre la base de realidad problemática presentada se planteó el problema general y los problemas específicos de la investigación. El problema general de la investigación fue ¿Cómo utilizar redes neuronales recurrentes para la predicción de casos por Covid-19 usando datos abiertos en el Perú, 2020?. Los problemas específicos de la investigación fueron los siguientes:

- **PE1:** ¿Cómo utilizar las redes neuronales recurrentes para predecir el aumento de la tasa de mortalidad de casos por Covid-19 usando datos abiertos en el Perú, 2020?

- **PE2:** ¿Cómo utilizar las redes neuronales recurrentes para predecir la variación de la tasa de letalidad de casos por Covid-19 usando datos abiertos en el Perú, 2020?

El objetivo general fue utilizar las redes neuronales recurrentes para predecir los casos por Covid-19 usando datos abiertos en el Perú, 2020. Los objetivos específicos fueron los siguientes:

- **OE1:** Utilizar las redes neuronales recurrentes para predecir el aumento de la tasa de mortalidad de casos por Covid-19 usando datos abiertos en el Perú, 2020
- **OE2:** Utilizar las redes neuronales recurrentes para predecir la variación de la tasa de letalidad de casos por Covid-19 usando datos abiertos en el Perú, 2020

La hipótesis general fue las redes neuronales recurrentes ayudan a predecir el aumento de los casos por Covid-19 usando los datos abiertos del Perú, 2020. Las hipótesis específicas fueron las siguientes:

- **HE1:** Las redes neuronales recurrentes ayudan a predecir el aumento de la tasa de mortalidad de Covid-19 usando los datos abiertos del Perú, 2020
- **HE2:** Las redes neuronales recurrentes ayudan a predecir la variación de la tasa de letalidad de Covid-19 usando los datos abiertos del Perú, 2020

II. MARCO TEÓRICO

En el año 2020, Moftakhar, L., Seif, M. y Safe, M., en su artículo científico titulada: “Tendencia exponencialmente creciente de pacientes infectados con covid-19 en Irán: una comparación de la red neuronal artificial (ANN) y los modelos de pronóstico de arima”, publicado por Iranian Journal of Public Health en Irán. El objetivo del artículo es pronosticar el número de casos diarios nuevos infectados con COVID-19 para los próximos treinta días en Irán. Se justifica, porque hace mención que la predicción de su estudio puede ser útil para la planificación de la salud y el gobierno. Para la población se han tomado los casos de infectados por Covid-19 en Irán y en la muestra solo se usaron los casos entre el 30 de febrero al 30 de marzo del 2020. El resultado obtenido de ambos algoritmos pronosticó un aumento exponencial en el número de pacientes recién infectados. Si el patrón de propagación sigue igual que antes, el número de nuevos casos diarios sería 7872 y 9558 en un 29^o Abr, respectivamente, por Ann y ARIMA. Mientras que la comparación de modelos confirmó que la predicción ARIMA fue más precisa que ANN. Por lo tanto, se llega a concluir que el COVID-19 es una enfermedad contagiosa y ha infectado a muchas personas en Irán. Nuestros resultados son una alarma para los planificadores de políticas de salud y los tomadores de decisiones, para tomar decisiones oportunas, controlar la enfermedad y proporcionar el equipo necesario. El componente reusable fue el error cuadrático medio y el error absoluto medio.

En el año 2020, Behrouz Pirouz, Tengo Shaffiee Haghshenas, Sami Shaffiee Haghshenas, Patrizia Piro, en su artículo científico titulada: “Investigando un serio desafío en el proceso de desarrollo sostenible: análisis de casos confirmados de COVID-19 (nuevo tipo de coronavirus) a través de una clasificación binaria usando inteligencia artificial y análisis de regresión”, fue desarrollada en la University of Calambria, Italia. El objetivo es analizar los casos confirmados de COVID-19 a través de una clasificación binaria usando inteligencia artificial y análisis de regresión. Para la población se utilizaron los casos confirmados de 42 provincias en China, Japón, Corea del Sur e Italia y para la muestra solo se utilizó los casos de la provincia de Hubei en China. Como resultado se obtuvo que las fluctuaciones diarias de casos confirmados con fluctuaciones de los parámetros climáticos en cuatro estudios de caso en China. La misma tendencia en la provincia de Hubei existe en las otras cuatro provincias, y la tasa aumentó

hasta el 4 de febrero y disminuyó con algunas fluctuaciones. En este estudio se investigó un serio desafío del desarrollo sostenible utilizando el algoritmo GMDH y el análisis de regresión. Según los resultados, el algoritmo GMDH tiene un rendimiento apropiado para predecir y clasificar los parámetros de un estudio de caso afectado por COVID-19, y las precisiones basadas en los conjuntos de datos de Wuhan fueron iguales a 95.7% y 85.7% para entrenamiento y pruebas, respectivamente. El componente reusable fue el análisis de los casos confirmados de Covid-19.

En el año 2020, Al-Najjar, H. y Al-Rousan, N., en su artículo científico titulada: “Un modelo clasificador para predecir el estado de los pacientes con Coronavirus CoVID-19 en Corea del Sur”, publicada por la European Review for Medical and Pharmacological Sciences, Turquía. Tiene como objetivo construir un modelo de predicción del clasificador para predecir el estado de pacientes con coronavirus CovID-19 recuperados y fallecidos en Corea del Sur. La población tomada fue los pacientes con coronavirus en Corea del Sur y se tomó como muestra a los Pacientes infectados entre el plazo del 29 de enero al 09 de abril del 2020. Los resultados encontraron que el clasificador predictivo propuesto predijo eficientemente los casos de recuperación y muerte. Además, se descubre que la razón de la infección aumentaría la probabilidad de recuperar al paciente. Esto indica que el virus puede ser controlable por razones de infección. Además, el descubrimiento anterior de la enfermedad permite un mejor control y una mayor probabilidad de ser recuperado. Se concluye que se implementó un modelo que puede clasificar con éxito si los pacientes actuales estarán en grupos recuperados o en grupos de muerte en función de varias variables. Su contribución reusable es modelo clasificador para poder predecir el estado de los pacientes.

En el año 2020, Duccio Fanelli y Francesco Piazza, en su artículo científico titulado: “Análisis y pronóstico de la propagación de COVID-19 en China, Italia y Francia”, publicado por Chaos, Solitons and Fractals, Italia. El objetivo es hacer una comparación de la evolución del brote de COVID-19 en China continental, Italia y Francia. Para la población se usaron casos por Covid-19 de Francia, Italia y China. En su resultado muestra que el modelo SIRD coloca el pico en Italia del 21 de marzo 2020, y predice un número máximo de individuos infectados

confirmados de aproximadamente 26.000 en el pico del brote. El número de muertes al final de las epidemias parece ser de aproximadamente 18000. Teniendo en cuenta que los casos confirmados pueden estimarse entre el 10 y el 20% del número real de individuos infectados, la tasa de mortalidad aparente de COVID -19 parece estar entre el 4% y el 8% en Italia, más alta que la gripe estacional, mientras que parece sustancialmente menor en China, es decir, entre el 1% y el 3%. Concluye que el análisis de los mismos datos dentro de un modelo simple de susceptibles infectados recuperados muertes (SIRD) revela que la tasa de recuperación es la misma para Italia y China, mientras que la tasa de infección y mortalidad parecen ser diferentes.

En el año 2020, Van-Dai Ta y CHUAN-MING Liu, en su artículo científico titulado: “Predicción de acciones basada en la optimización de la cartera utilizando una red de memoria a largo plazo en operaciones cuantitativas”, desarrollada en la Universidad Nacional de Tecnología de Taipei, Taiwán. Su objetivo es proponer la construcción de un modelo para la predicción del precio de las acciones, donde luego se aplique técnicas de optimización de la cartera para aprovechar los resultados de la predicción Múltiple. Propone una investigación de enfoque cuantitativo. Como resultado, podemos aprovechar la predicción para construir una cartera cuantitativa para cada horizonte de tiempo previsto. Sobre la base de las técnicas de optimización, nuestras carteras construidas se desempeñaron de manera efectiva al obtener altos rendimientos tanto de predicción como de negociación real, así como en comparación con el índice S & P 500. Este artículo presenta la red LSTM, un tipo de red neuronal recurrente, para predecir el precio de las acciones con el fin de demostrar un valor cuantitativo típico estrategia comercial. El modelo propuesto funciona de manera eficiente al lograr una alta precisión en comparación con otro aprendizaje automático como LR y SVM. El componente reusable fue la construcción del modelo predictivo.

En el año 2020, Altamirano Pontigo, Sergio Esteban, en su tesis titulada: "Aplicación de redes neuronales recurrentes y modelos de series de tiempo bayesianos a la predicción de rentabilidad de fondos de pensiones", para optar el grado de Para optar al título profesional de Ingeniero Estadístico, desarrollada en la Universidad de Concepción, Chile. El objetivo es desarrollar un estudio

comparativo sobre la aplicación de técnicas de predicción para series de tiempo a la rentabilidad de los fondos de pensiones. Para el caso del modelo LSTM, considerando al Fondo E, el máximo RMSE obtenido es de 1956,6 y un mínimo de solo 39,4, alcanzando un promedio de 214,6 y una desviación estándar de 212,1. Además, se obtiene en promedio un MAPE menor al Fondo A, pues es solo de un 0,5 con una desviación de 0,5. Al observar el sentido de la tendencia, en promedio se obtuvo un sentido de 73,5 %, 2 unidades porcentuales más que el Fondo A, con una desviación estándar de 18,7 %. si se compara entre herramientas, el modelo LSTM en la gran mayoría de los casos obtuvo mejores resultados que Prophet en el sentido de la tendencia, pero, hubo un caso único y particular, donde Prophet obtuvo mejor resultado, específicamente en el Fondo E, donde alcanzó un 62 % en el Sentido, siendo el único caso que superó los 60 %. Entonces definir qué herramienta obtiene los mejores resultados, debe definirse por otras métricas.

En el año 2020, Jorge Puentes Márquez, Celma de Oliveira Ribeiro, Edgar Ruelas Santoyo, Vicente Figueroa Fernández, en su artículo científico titulada: Pronóstico de la demanda de combustible de etanol en Brasil uso de una red neuronal recurrente LSTM, desarrollada en la IEEE Latin America Transactions, Brasil. El objetivo es pronosticar el consumo de etanol en Brasil usando un enfoque de modelos de Redes Neuronales Recurrentes (RNN) de Memoria a Largo-Corto Plazo (LSTM) y Promedio Móvil Integrado Autoregresivo (ARIMA), la serie de tiempo no es estacionaria y que su media y varianza cambian a través del tiempo. Esto da indicios de que se posee tendencia o estacionalidad en dicha serie. Los modelos LSTM encontrados mediante el algoritmo de búsqueda de topología de red, tanto univariados como multivariados, tuvieron el mejor desempeño para el caso de estudio abordado.

En el año 2018, Elena Rosa Pérez, Federico Liberatore, en su tesis titulada: "Predicción demanda eléctrica española. Implementación de redes neuronales recurrentes en Python", para optar el grado de máster en Ingeniería Matemática, desarrollada en la Universidad Complutense de Madrid, España. El objetivo es la predicción de la demanda eléctrica nacional mediante el desarrollo de modelos de redes neuronales recurrentes. Como vemos obtenemos un peor resultado al

predecir directamente a nivel horario, tanto con la célula LSTM como con la GRU, que predecir a nivel horario y luego desagregar. De hecho, este error ha aumentado en más de un 1%. Se han estudiado de forma teórica las redes neuronales y cómo funcionan (en especial las recurrentes). Además, se ha realizado un análisis de la demanda eléctrica nacional y se han comparado diferentes metodologías de predicción.

En el año 2020, Shastri, Singh, Kumar, Kour y Mansotra, en su artículo científico titulado: "Pronóstico de series de tiempo de Covid-19 utilizando modelos de aprendizaje profundo: estudio de caso comparativo India-EE. UU.", desarrollada en la Universidad de Jammu, India. El objetivo es proponer modelos basados en aprendizaje profundo para la predicción de los casos confirmados y de muerte de Covid-19 para India y EE. UU. En sus resultados presenta una comparación de 3 modelos (Stacked LSTM, Bi-directional LSTM y Convolutional LSTM), los cuales se utilizó para modelar las predicciones y en base al error de predicción (MAPE) determinar que modelo es más preciso. En su conclusión indica que el modelo ConvLSTM supera a los modelos Stacked LSTM y Bi-directional LSTM. El componente reusable de esta investigación es el modelo Bi-directional LSTM.

En el año 2020, Tölö, E., en su artículo científico titulado: "Predecir crisis financieras sistémicas con redes neuronales recurrentes", desarrollada en la University of Helsinki, Finlandia. El objetivo es predecir eventos de crisis financiera sistémica utilizando redes neuronales recurrentes. En sus resultados evaluó el rendimiento de predicción de los modelos de redes neuronales, comparándolos con un modelo logístico con un solo rezago y múltiples (5) rezagos, denominados Logit (1) y Logit (5), respectivamente. Concluye que el uso de la entrada de series de tiempo puede conducir a predicciones más precisas y que las RNN, especialmente los RNN cerrados (RNN-LSTM y RNN-GRU), superan al modelo logit y a las redes neuronales perceptrón multicapa.

En el año 2017, Rodríguez Paredes Noelia Patricia, en su tesis titulada: "Pronóstico del caudal máximo del Río Huaura por modelos Arima y redes neuronales", para optar el grado de título profesional de Ingeniero Estadístico, desarrollada en la Universidad Nacional de Trujillo, Lima. El objetivo es determinar la

eficiencia de los modelos arima en el pronóstico del caudal máximo del río Huaura en comparación con las redes neuronales. Como justificación indica que aporta para evitar ocurrencias de eventos máximos de precipitación, ya que generan problemas. Metodología de desarrollo fue Modelos arima y redes neuronales y propone una investigación de tipo aplicada con enfoque cuantitativo, su diseño es no experimental. Para la población se tomaron los caudales máximos mensuales del río Huaura desde su existencia hasta el mes de Julio del 2016 y en su muestra se utilizó los caudales máximos mensuales del periodo hidrológico de enero de 1960 a julio del 2016. Se tiene de resultado que los caudales máximos del río Huaura se analizaron con un total de 679 datos para realizar la estimación del modelo. Se puede indicar que el comportamiento de los caudales máximos del río Huaura es asimétrica positiva a la derecha con un coeficiente de asimetría de 1,36 y es leptocúrtica porque su coeficiente de curtosis es mayor a 3. Se concluye que los modelos de redes neuronales son más eficientes que los modelos arima al obtener los menores errores de pronóstico. El componente reusable fue su metodología de investigación.

En el año 2020, Anibal Flores, Hugo Tito y Deymor Centty, en su artículo científico titulado: "Comparación redes neuronales recurrentes híbridas para la predicción de series de tiempo univariadas", desarrollada en la Universidad Nacional de Moquegua, Perú. Tiene como objetivo mejorar la precisión de los modelos de pronóstico de series de tiempo utilizando modelos híbridos de redes neuronales recurrentes. En sus resultados realiza una comparación de 2 modelos no híbridos LSTM y GRU con 14 modelos híbridos LSTM-GRU realizando todas las combinaciones posibles. De estos modelos se muestra el RMSE respectivo para cada caso y así obtener el más confiable. Para arquitecturas de 4 capas de 14 casos de estudio, las arquitecturas híbridas superan a las arquitecturas no híbridas en el 50% de los casos. Generalizando desde 28 estudios de caso en 21 de ellos, las arquitecturas híbridas superan a las no híbridas. Por lo tanto, se concluye que, para las series temporales analizadas en este estudio, arquitecturas híbridas basadas en redes neuronales recurrentes son una gran alternativa para el pronóstico de series de tiempo univariantes. El componente reusable de esta investigación fue el modelo híbrido y la técnica de cómo obtener el mejor modelo para una predicción de series de tiempo.

En el año 2019, Percy Cueva Rios, en su tesis titulada “Evaluación de índices de confiabilidad en redes de Electrocentro S.A, para mejorar la calidad de servicio utilizando redes neuronales”, para optar el grado de Doctor en Ingeniería Eléctrica, desarrollada en la Universidad Nacional del Callao, Perú. El objetivo es implementar un método de diagnóstico de los índices de confiabilidad basados en el SAIDI y SAIFI y realizar la mejora de la calidad de servicio en las redes de Electrocentro S.A. Su justificación planteó una propuesta de evaluación de los índices de confiabilidad, basados en el software Digsilent, pero para la realidad actual y con los considerandos establecidos en la Norma Técnica de Calidad de Servicio Eléctrico. Su metodología a investigación es de tipo tecnológico con un enfoque cuantitativo. Se tomó como población a los reportes de interrupción de servicio y duración de los mismos en las redes de Electrocentro S.A. La muestra a considerar, fueron los alimentadores de la Sub Estación NinaTambo A4701 y A4702. Se obtiene como resultado de las necesidades de la empresa concesionaria Electrocentro S.A. sería el que más se adecua a su capacidad de inversión, esto teniendo en consideración que en muchos lugares los índices de morosidad son bastante elevados y no justifican realizar grandes inversiones que no resultan rentables para el concesionario. Y su conclusión es que un modelamiento de las redes, en forma muy particular, nos permitiría una reducción de los índices de confiabilidad a valores próximos a los recomendados en la Resolución Osinergmin N° 178- 2012-OS/CD y que permitan su viabilidad de implementación.

En el año 2019, Edwin Rober Córdova Delgado, Jhojan Pierre Mautino Vidaurre, en su tesis titulada: “Análisis de redes neuronales artificiales para el pronóstico de las enfermedades prevalentes en el hospital Víctor Ramos Guardia de Huaraz, 2016”, para optar el grado de Licenciado en Estadística e Informática, desarrollada en la Universidad Nacional Santiago Antúnez de Mayolo, Perú. El objetivo es pronosticar las enfermedades prevalentes mediante el uso de las redes neuronales artificiales. Ante ello los modelos matemáticos y estadísticos, son los que permitirán representar lo que sucede en el diagnóstico de la enfermedad, frente a los distintos patógenos, los cuales se desarrollan dependiendo de los objetivos del trabajo, los datos disponibles, el tipo de experimento y la experiencia del investigador. Metodología de desarrollo fue y propone una investigación

tipo aplicada, enfoque cuantitativo, diseño no experimental y longitudinal. Su población son los registros de atenciones del hospital Víctor Ramos Guardia – Huaraz, hasta el 2016 y su muestra fueron los registros del 2011 al 2016. Las enfermedades prevalentes en el Hospital, son caracterizados por enfermedades categorizados como factores que influyen en el estado de salud y contacto con los servicios de salud (E1: 24,3%), traumatismos, envenenamientos y algunas otras consecuencias de causa externa (E2: 21,0%), enfermedades del sistema respiratorio (E3: 12,9%), síntomas, signos y hallazgos anormales clínicos y de laboratorio, no clasificados en otra parte (E4: 12,4%) y ciertas enfermedades infecciosas y parasitarias (E5: 8,5%). El pronóstico de las Redes Neuronales Artificiales es adecuado y con convergencia del error de estimación a cero, sin superar el umbral o límite de error de estimación del 1% con un tope máximo de 1000000 iteraciones. El componente reusable fue su metodología de investigación.

En el año 2020, Anibal Flores, Hugo Tito and Deymor Centty, en el artículo titulada “Redes neuronales recurrentes para tiempo meteorológico Imputación de serie”, desarrollada en la Universidad Nacional de Moquegua, Perú, El objetivo es analizar la efectividad de las redes neuronales recurrentes en procesos de imputación de series de tiempo meteorológicas, para esté diferentes modelos basados en redes neuronales recurrentes cómo Memoria a corto plazo (LSTM) y unidad recurrente con compuerta (GRU) se implementan y se experimenta con horas series de tiempo meteorológico como temperatura, dirección del viento y velocidad del viento. De acuerdo con lo que se muestra en la Tabla III (resultados de temperatura de series de tiempo) y en la Fig.11 (Los 3 mejores modelos RNN para la imputación de series temporales de temperatura) para el proceso de imputación en la serie de tiempo de temperatura en un promedio, el mejor modelo es LSTM LSTM LSTM (RMSE 0.5565), este modelo también fue el que produjo el mejor resultado para todos los tamaños de huecos. Los resultados obtenidos muestran que no todos los modelos logran resultados óptimos, por lo que es importante implementar no solo un modelo, pero varios de tal manera que el más apropiado Se puede elegir el modelo para resolver el problema. En las tres series de tiempo analizadas, los modelos basados en LSTM muestran mayor heterogeneidad en sus resultados en comparación con GRU-modelos basados

en resultados más homogéneos. El componente reusable es el modelo de red neuronal recurrente.

En el año 2017, Grandez, en su tesis titulada: Aplicación de minería de datos para determinar patrones de consumo futuro en clientes de una distribuidora de suplementos nutricionales, para optar el grado de Título Profesional de Ingeniero Informático y de sistemas., desarrollada en la Universidad San Ignacio de Loyola, Perú. Tiene como objetivo identificar a los compradores potenciales basándose en sus características demográficas e historial de transacciones, de acuerdo a los clics que realiza. Metodología de desarrollo fue CRISP-DM y propone una investigación Tipo de investigación: Exploratorio Diseño de investigación: No experimental de tipo transversal. Su población son los datos de clientes de la distribuidora de suplementos nutricionales-local Miraflores. Para la muestra utilizó datos de 611 clientes que realizaron compras entre enero del 2016 a junio del 2016. En sus resultados coloca las probabilidades de que un cliente compre un producto. Y concluye que la minería de datos es un área de conocimiento que nos ayuda a tomar decisiones en base a información del propio negocio. El componente reusable de esta investigación es la metodología CRISP-DM.

Para entender lo que es una red neuronal primero debemos entender cuál es su origen, y para ello se menciona que “las redes neuronales se inspiran en la manera en cómo trabaja el cerebro humano. [...] puede procesar grandes cantidades de información utilizando los datos enviados por los sentidos humanos (especialmente la visión)” (Ciaburro y Venkateswaran, 2017). Podemos inferir que las redes neuronales buscan modelar el funcionamiento que realizan las neuronas en el cerebro humano, ya que tienen la capacidad de hacer grandes procesamiento de información.

La red neuronal se define como “un sistema que posibilita construir una relación entre entradas y salidas, se inspira en el sistema nervioso y distinguiéndose de la computación tradicional, debido a que estos no hacen uso de un algoritmo secuencial” (Serna, 2017). Por otra parte, se indica que “[...] una red neuronal es un algoritmo que se compone de algoritmos diferentes, que realizan cálculos locales más pequeños a medida que los datos se propagan a través de

ella. [...] Una red neuronal es un mecanismo que toma entradas y aprende asociaciones para predecir algunas salidas” (Purkait, 2019).

La red neuronal artificial o ANN (Artificial Neural Networks) “es una herramienta de modelado de datos que depende de varios parámetros y métodos de aprendizaje [...]. ANN adquiere conocimiento a través del aprendizaje, y este conocimiento se almacena dentro de la fuerza de las conexiones interneuronas, que se expresa mediante valores numéricos llamados pesos” (Chakraverty y Mall, 2017). También se menciona que “la red neuronal artificial se utiliza en muchos campos diferentes para procesar grandes conjuntos de datos, a menudo proporcionando análisis útiles que permiten la predicción e identificación de nuevos datos” (El-shahat, 2018). El siguiente autor indica que la ANN “están diseñadas de manera que puedan identificar los patrones subyacentes en los datos y aprender de ellos. Se pueden utilizar para diversas tareas, como clasificación, regresión, segmentación, etc. Necesitamos convertir cualquier dato dado en forma numérica antes de alimentarlo a la red neuronal” (Joshi, 2017a).

Ahora definiremos el tipo de aprendizaje de las redes neuronales. “El aprendizaje supervisado es el enfoque más utilizado y mejor entendido para el aprendizaje automático. Implica una entrada y salida para cada pieza de datos en su conjunto de datos” (Norman, 2017). Otro autor menciona que el aprendizaje supervisado “[...] realiza una comparación entre la salida calculada de la red y la salida esperada corregida para determinar el error. El error se puede usar para cambiar los parámetros de red, dando así al rendimiento una mejoría” (Chakraverty y Mall, 2017). Y por último indica que, [...] un modelo de aprendizaje automático que se basa en datos de capacitación etiquetados. Por ejemplo, supongamos que queremos construir un sistema para predecir automáticamente los ingresos de una persona, en función de diversos parámetros, como la edad, la educación, la ubicación, etc. Para hacer esto, necesitamos crear una base de datos de personas con todos los detalles necesarios y etiquetarla. Al hacer esto, le estamos diciendo a nuestro algoritmo qué parámetros corresponden a qué ingreso. Con base en este mapeo, el algoritmo aprenderá cómo calcular el ingreso de una persona utilizando los parámetros que se le proporcionan (Joshi, 2017b).

El aprendizaje no supervisado es donde “los datos de entrenamiento no incorporan las etiquetas y será el algoritmo el que intentará clasificar la información por sí mismo” (Torres, 2018). Por otra parte, el aprendizaje no supervisado “[...] generalmente tiene como objetivo descubrir las similitudes y patrones entre las muestras o reproducir una distribución de entrada dado un conjunto de vectores extraídos de ella” (Bonaccorso, 2019). Para poder entender mejor lo que es el aprendizaje no supervisado, [...] un modelo de aprendizaje automático sin depender de datos de entrenamiento etiquetados. [...] Como no hay etiquetas disponibles, debe extraer información basada solo en los datos que se le proporcionaron. Por ejemplo, supongamos que queremos construir un sistema donde tengamos que separar un conjunto de puntos de datos en múltiples grupos. Lo complicado aquí es que no sabemos exactamente cuáles deberían ser los criterios de separación. Por lo tanto, un algoritmo de aprendizaje no supervisado necesita separar el conjunto de datos dado en varios grupos de la mejor manera posible (Joshi, 2017c).

En esta parte conoceremos al perceptrón simple donde indica que “[...] tienen nodos de entrada y salida y una sola capa de conexiones entre ellos. [...] Un perceptrón de una sola capa tendría éxito en una tarea de aprendizaje si existiera una solución a la tarea. La importancia de este resultado radica en el hecho de que el aprendizaje podría comenzar desde diferentes lugares y con diferentes ejemplos” (Wells, 2017). Para una segunda definición, “un perceptrón es el bloque de construcción de una red neuronal artificial. Viene a ser una única neurona que hace uso de entradas, elabora cálculos sobre ellas para luego generar una salida. Hace uso de la función lineal simple para haga la decisión” (Joshi, 2017d).

“Un perceptrón multicapa es una red en la que las neuronas se agrupan en capas. Dicha red tiene una capa de entrada, una o más capas ocultas y una capa de salida (Patan, 2019). Por otro parte el perceptrón multicapa “puede aceptar múltiples entradas a la vez por predicción. Esto permite a la red combinar diversas formas de información para tomar decisiones mejor informadas” (Trask, 2019b).

La red neuronal recurrente “[...] (también llamada feedback) utiliza un estado de memoria adicional. Cuando se agrega una entrada A1 (word I), la red

produce una salida B1 (word love) y almacena información sobre la entrada A1 en el estado de memoria. Cuando se agrega la siguiente entrada A2 (word love), la red produce la salida asociada B2 (word to) con la ayuda del estado de la memoria. Luego, el estado de la memoria se actualiza utilizando la información de la nueva entrada A2” (Kostadinov, 2018).

Las redes recurrentes simples (SRN) “rara vez se usan en las aplicaciones, pero son el principal método de enseñanza para explicar las redes recurrentes antes de ejecutar en el LSTM más complejo, que son la arquitectura recurrente principal que se usa hoy en día” (Skansi, 2018).

“El SRN combina la clásica red de retropropagación de tres capas con una capa recurrente de unidades de contexto, que pueden mantener una copia de las activaciones de unidades ocultas en un momento anterior” (Hundt, Mollin y E. Pfenninger, 2017)

“El LSTM tiene conexiones recurrentes para que el estado de las activaciones previas de la neurona del paso de tiempo anterior se use como contexto para formular una salida. Pero a diferencia de otros RNN, el LSTM tiene una formulación única que le permite evitar los problemas que impiden el entrenamiento y la ampliación de otros RNN” (Brownlee, 2019).

Los LSTM crean el siguiente estado oculto copiando el estado oculto anterior y luego agregar o eliminar información según sea necesario. Los mecanismos que utiliza LSTM para agregar y eliminar información se llaman puertas. [...] LSTM almacena información en una serie de tiempo sin preocuparse sobre la desaparición o explosión de gradientes (Trask, 2019a).

Gated Recurrent Unit (GRU), “Esta red funciona de la misma manera que un modelo RNN normal en términos de entradas, estados de memoria y salidas. La diferencia clave radica en los detalles de la celda en cada paso de tiempo” (Kostadinov, 2018).

El GRU “consta de dos puertas: una puerta de actualización y reinicio. La actualización cerrada está diseñada para ayudar a aprender dependencias a

largo plazo de una serie de entradas, mientras que la puerta de reinicio está destinada a aprender dependencias a corto plazo” (Agarwal, 2020).

El coronavirus forma parte de la familia Coronaviridae, que viene a ser sub-familia Coronavirinae. Son virus encapsulados con un diámetro entre 60-140 nm, tienen un ARN no segmentado, positivo y largo (27,9 kb y 30,1 kb). Se clasifican en 4 géneros que son Alfacoronavirus, Betacoronavirus, Deltacoronavirus y Gamacoronavirus; solo algunas especies pertenecientes a los dos primeros han sido descritas como agentes etiológicos de enfermedades respiratorias en humanos (adultos y niños) (Quiroz *et al.*, 2020).

El reciente brote de COVID-19 se ha considerado una emergencia sanitaria mundial. A nivel internacional, el número de informes confirmados ha seguido aumentando, y actualmente se ubica en 90.870 casos confirmados por laboratorio con más de 3.000 muertes. Quizás esté claro que la cuarentena por sí sola puede no ser suficiente para la prevención de la propagación de COVID-19, y el impacto global de esta infección viral es motivo de gran preocupación. (Sohrabi *et al.*, 2020).

La serie de tiempo es “una variable bidimensional (t , Y_t), en la que una de las componentes, la dependiente, es la magnitud que queremos analizar, mientras que la independiente es el tiempo” (Juan García *et al.* 2007, citado por Neira, 2017)

“Una de las principales razones de estudio de las series de tiempo, es la generación de pronósticos. Como las series analizadas son estocásticas, su comportamiento futuro se puede determinar de forma parcial haciendo uso de las distribuciones de probabilidad generadas por sus valores pasados.” (Mota Arantxa, 2016)

“Si los datos históricos se rigen a valores pasados de la variable que tratamos de pronosticar, el procedimiento de elaboración de pronósticos se llama método de serie de tiempo. El objetivo de los métodos de serie de tiempo es descubrir patrones en los datos históricos y luego extrapolarlos hacia el futuro” (Anderson, Sweeney, Williams, Camm, & Martin, 2011, citado por Ruiz, 2018).

“La predicción de series de tiempo considera la hipótesis que dado un conjunto discreto de datos con respecto al tiempo y correspondientes a un mismo fenómeno, los valores futuros son dependientes de los valores pasados, es decir, buscando en los valores pasados de una serie de tiempo se puede predecir su comportamiento en el futuro” (Escobar et al. 2009, citado por Neira Mike, 2017).

“Métricas de predicción. Tres términos de El error predictivo se utilizan para la evaluación experimental. El primero es el porcentaje de error absoluto (abreviatura APE) definido como la ecuación. 1; el segundo es el error porcentual absoluto medio (abreviado MAPE) como la ecuación. 2; el tercero es la mediana error porcentual absoluto (abreviatura MDAPE) como la ecuación 3)” (Ding *et al.*, 2020).

Dimensión: Predicción de casos

La predicción de casos servirá para “proporcionar a los funcionarios estimaciones realistas del tiempo y la magnitud del pico epidémico, es decir, el número máximo de personas infectadas [...]. Los datos se obtuvieron del repositorio de GitHub asociado con el tablero interactivo alojado por el Centro de Ciencia e Ingeniería de Sistemas (CSSE) en la Universidad Johns Hopkins” (Fanelli y Piazza, 2020).

Indicadores:

Tasa de mortalidad (TM). - “La tasa de mortalidad general es la proporción de personas que fallecen respecto al total de la población (usualmente expresada en tanto por mil, 0/00), la tasa de mortalidad particular se refiere a la proporción de personas con una característica particular que mueren respecto al total de personas que tienen esa característica” (Asencios, 2015) Utiliza la siguiente fórmula: $TM = \frac{TF}{NH} \times 10000$. En donde TF = Total de fallecidos, NH = Número de habitantes en el Perú (32 625 948 según Instituto Nacional de Estadística e Informática (INEI)).

Tasa de letalidad (TL). - “Finalmente, estamos interesados en estimar la tasa de letalidad que mide qué tan letal es el COVID-19 para las personas infectadas” (He, Yi y Zhu, 2020). También, “la letalidad es una medida de la gravedad de una enfermedad considerada desde el punto de vista poblacional, y se define

como la proporción de casos de una enfermedad que resultan mortales con respecto al total de casos en un periodo especificado. La medida indica la importancia de la enfermedad en términos de su capacidad para producir la muerte” (Moreno, López y Corcho, 2000). Utilizar la siguiente fórmula: $TL = \frac{TF}{TC} \times 100$. En donde TL = Tasa de letalidad, TF = Total de fallecidos por Covid-19 y TC = Total de casos confirmados por Covid-19.

Considerando el estado del arte de trabajos previos de las investigaciones más destacadas, los autores no utilizaron una metodología para el desarrollo de las redes neuronales artificiales a excepción de (Rodríguez, 2017), quien utilizó la metodología Box-Jenkins, el resto se ha basado en arquitecturas para la elaboración de su proyecto. En nuestra investigación se ha tomado en cuenta las metodologías de desarrollo para sistemas ágiles, en la cual se distinguen 3 metodologías que brindan los pasos para cumplir con el desarrollo del proyecto, de las cuales se eligió uno de acuerdo a los resultados del juicio de expertos. La metodología XP obtuvo un puntaje de 24%, KANBAN 27% y SCRUM 49%. Finalmente, la metodología a usar fue SCRUM, el cual admite fases que permiten el desarrollo del sistema. SCRUM es una metodología ágil que utilizan los desarrolladores de software que a menudo tienen que abordar proyectos grandes y complejos, proporciona conocimiento corporativo e incluye métodos que realizan un análisis detallado de tareas y procesos intensivos en conocimiento. (Jessica Keyes, 2016).

Prosiguiendo con el manejo de los datos se optó trabajar con una metodología de minería de datos de la cual al momento de elegir entre la metodología CRISP-DM, SEMMA y KDD. Se tomó la decisión de implementar la Metodología CRISP-DM.

TABLA 1 CUADRO COMPARATIVO METODOLOGÍAS DE MINERÍA DE DATOS

CRITERIOS/METODOLOGÍAS	CRISP-DM	SEMMA	KDD
Metodología Estructurada	✓	✓	✓
Metodología Independiente	✓	X	✓
Ampliamente Usada	✓	X	X
Mejora la calidad de resultados en proyectos de Data Mining.	✓	✓	✓
Herramientas y técnicas independientes	✓	✓	✓
Finalidad diversa (Ej. Ampliamente estable en la resolución de problemas variados).	✓	✓	✓
Fácil de implementar	✓	✓	✓

Fuente: Grández Márquez Miguel

III. METODOLOGÍA

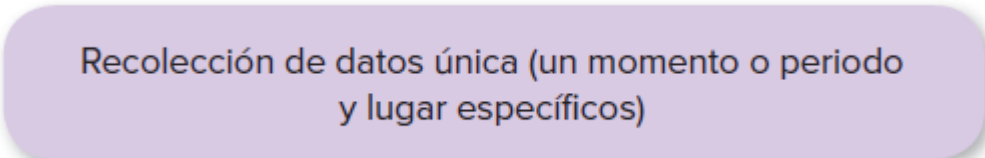
3.1 Tipo y diseño de investigación

La presente investigación será de tipo aplicada, porque buscará dar una respuesta a los problemas planteados en el capítulo anterior. La investigación de tipo aplicada es un “tipo de investigación orientada a la búsqueda de nuevos conocimientos sin una finalidad práctica específica e inmediata” (Sánchez, Reyes y Mejía, 2018).

La presente investigación será de enfoque cuantitativo porque “[...] representa un conjunto de procesos organizado de manera secuencial para comprobar ciertas suposiciones. [...] Parte de una idea que se delimita y, una vez acotada, se generan objetivos y preguntas de investigación, se revisa la literatura y se construye un marco o perspectiva teórica.” (Hernandez y Mendoza, 2018).

Dado que el objetivo del estudio será predecir el número de casos de las personas infectadas y fallecidas a causa del Covid-19 en el Perú, se recurrirá a un diseño no experimental que se aplicará de manera transversal. De acuerdo con el diseño no experimental es una “investigación que se elabora sin alteración de variables, lo que se hace, es contemplar fenómenos tal como ocurren, para continuar analizando posteriormente” (Escobar y Bilbao, 2020). Además, también indican que en el diseño transversal se “[...] recolectan datos en un solo momento, en un tiempo único [...] Evaluar una situación, comunidad, evento, fenómeno o contexto en un punto del tiempo [...]. Analizar la incidencia de determinadas variables, así como su interrelación en un momento, lapso o periodo” (Hernandez y Mendoza, 2018).

Figura 1: Diseño no experimental - transversal

El diagrama consiste en un único recuadro rectangular con esquinas redondeadas, de color lavanda. El recuadro está centrado horizontalmente y contiene el texto "Recolección de datos única (un momento o periodo y lugar específicos)".

Recolección de datos única (un momento o periodo
y lugar específicos)

Fuente: Hernandez y Mendoza

3.2 Variables y operacionalización

a) Definición Conceptual:

Variable independiente: Redes neuronales recurrentes

En particular, las redes neuronales recurrentes son excelentes para modelar datos secuenciales. [...] Cuando trabajamos con datos de series de tiempo, no podemos simplemente usar modelos de aprendizaje genéricos. Necesitamos caracterizar las dependencias temporales en nuestros datos para que podamos construir un modelo robusto. (Joshi, 2017e)

Variable dependiente: Predicción de casos por Covid-19

La predicción de casos servirá para “proporcionar a los funcionarios estimaciones realistas del tiempo y la magnitud del pico epidémico, es decir, el número máximo de personas infectadas [...]. Los datos se obtuvieron del repositorio de GitHub asociado con el tablero interactivo alojado por el Centro de Ciencia e Ingeniería de Sistemas (CSSE) en la Universidad Johns Hopkins” (Fanelli y Piazza, 2020).

b) Definición Operacional:

Variable independiente: Redes neuronales recurrentes

La variable se utilizará de apoyo para obtener los resultados de la predicción de 7 días a futuro. (Joshi, 2017f).

Variable dependiente: Predicción de casos por Covid-19

La variable se analiza en función a dos indicadores los cuales son la tasa de mortalidad del cual es conocer el cociente entre el número total de muertes ocurridas en un país en un periodo determinado y el número de habitantes de la población de referencia y la tasa de letalidad para conocer el número de personas fallecidas por una determinada enfermedad en un periodo de tiempo determinado (Fanelli y Piazza, 2020).

3.3 Población, muestra y muestreo

Se explica que “El número de elementos experimentales que conforman nuestra población determinará si trabajar con ella en su totalidad o con una muestra representativa” (Cañadas y San Luis, 2018).

3.3.1 Población

“La población es un conjunto de elementos que contienen ciertas características que se pretenden estudiar. [...] Las poblaciones deben situarse de manera concreta por sus características de contenido, lugar y tiempo, así como accesibilidad” (Hernandez y Mendoza, 2018). La población a considerar es de 963605 registros de confirmados y 35966 registros de fallecidos por Covid-19 encontrados en la Plataforma de Datos Abiertos del Perú, hasta el 29/11/2020 estratificadas en días.

Criterios de inclusión: Para realizar la investigación se tomarán en cuenta los registros del número de casos confirmados y fallecidos por Covid-19.

Criterios de exclusión: No se tomarán en cuenta los números de casos de recuperados y ni la cantidad de pruebas realizadas durante el Covid-19.

Delimitación geográfica: Es el estudio se realizará con los casos registrados por covid-19 en el Perú.

3.3.2 Muestra

Se indica que “[...] es el grupo de personas, eventos, sucesos, comunidades, etc., sobre el cual se habrán de recolectar los datos, sin que necesariamente sea estadísticamente representativo del universo o población que se estudia” (Hernández y Mendoza, 2018). Para la muestra es de 7 registros ya estratificados desde el 23/11/2020 – 29/11/2020.

3.3.3 Muestreo

La investigación no probabilística “[...] también denominadas muestras dirigidas, como ya te mencionamos, suponen un procedimiento de selección orientado por las características y contexto de la investigación, más que por un criterio estadístico de generalización.” (Hernández y Mendoza, 2018)

Por la naturalidad de la investigación se requiere hacer uso de todos los datos por consiguiente se optó utilizar el muestreo por conveniencia. Indica que el muestreo por conveniencia “consiste en seleccionar las unidades muestrales más convenientes para el estudio o en permitir que la participación de la muestra sea totalmente voluntaria. Por lo tanto, no existe control de la composición de la

muestra y la representatividad de los resultados es cuestionable” (Fernández, 2004).

Unidad de análisis: Registro de persona por caso confirmado Covid-19 en el Perú.

Unidad de análisis: Registro de persona por caso fallecido Covid-19 en el Perú.

3.4 Técnicas e instrumentos de recolección de datos

3.4.1 Técnica

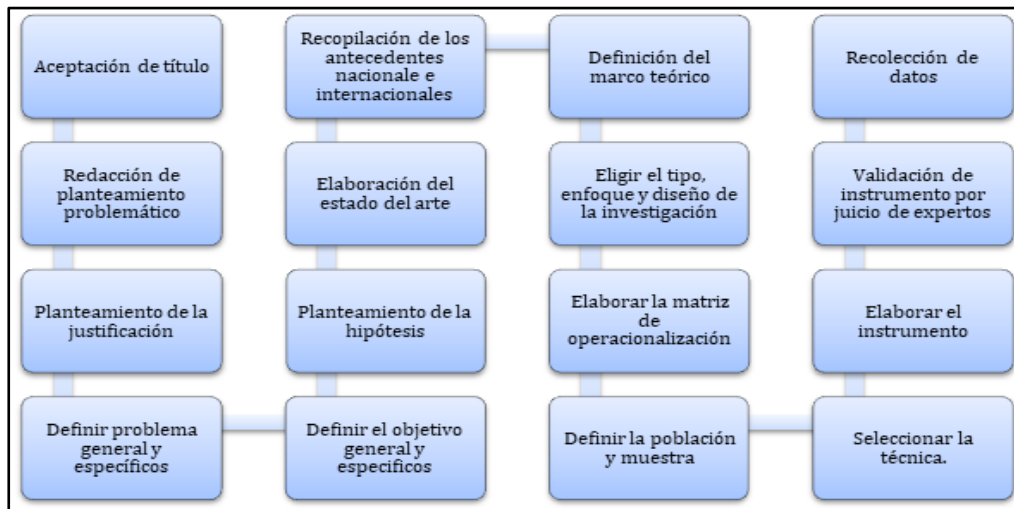
“El fichaje es la técnica que permite el registro de información seleccionada para el proceso de investigación. Su aplicación requiere el uso de fichas para ayudarnos a recoger y organizar la información” (Parraguez, Chunga, Flores y Romero, 2017). La técnica que se aplicará será el fichaje porque se va seleccionar solo la información necesaria de cada registro.

3.4.2. Instrumento

Define que la ficha de registro es “instrumento de recolección de la información[...] aplicada a libros, boletines, revistas, folletos, y periódicos que se utilizaron como fuentes para recolectar datos sobre las categorías de interés” (Zambrano, O., Espitia, S., & Hernández, J. 2017). El instrumento que se utilizará es la ficha porque es un instrumento ideal para poder organizar los datos que se recopilan de los registros de la base de datos.

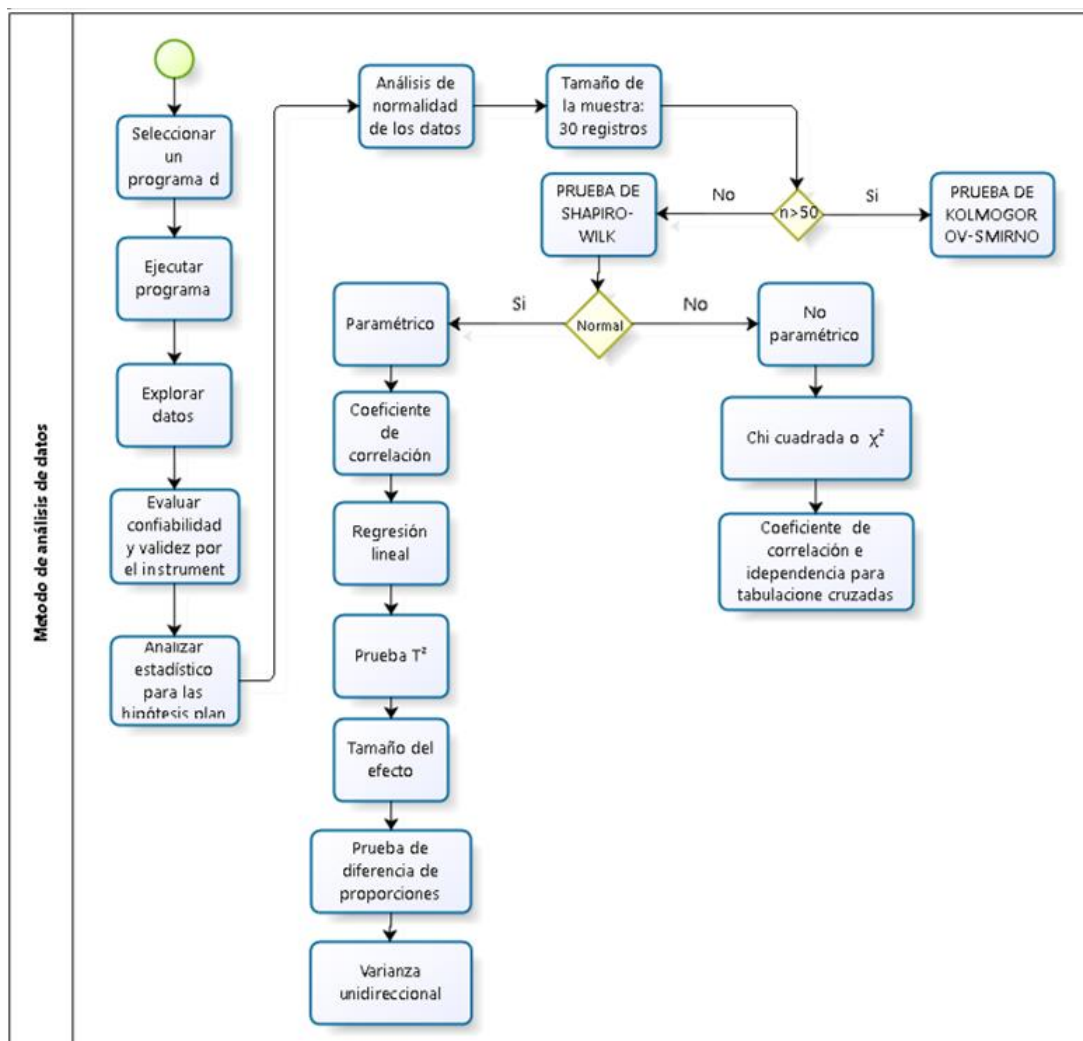
3.5 Procedimientos

Figura 2: Procedimiento de proyecto



3.5 Método de análisis de datos

Figura 3: Método de análisis de datos



3.7 Aspectos éticos

Para el desarrollo de la presente investigación se utilizará información de las personas afectadas por Covid-19 que fueron publicadas por el gobierno peruano, la cual contiene información que no comprometa o afecte a su integridad (física y psicológica). La investigación respeta la autoría de la información de otros autores, utilizando citas de manera correcta, la cual se desarrolló basándose en la Norma ISO 690.

IV. RESULTADOS

El presente capítulo detalla los resultados de la investigación teniendo en cuenta los indicadores de “Tasa de mortalidad” y “Tasa de letalidad”. Como también se podrá observar el procesamiento de la información de la elaboración del modelo para realizar la predicción de los casos por Covid-19.

4.1 Contraste de hipótesis

Indicador 1: Aumento de la tasa de mortalidad por cada 10 000 habitantes

HIPÓTESIS ESPECÍFICA

Las redes neuronales recurrentes ayudan a predecir el aumento de la tasa de mortalidad de Covid-19 usando los datos abiertos del Perú, 2020

HIPÓTESIS ESTADÍSTICAS

Hipótesis Nula (H0):

Las redes neuronales recurrentes no ayudan a predecir el aumento de la tasa de mortalidad de Covid-19 usando los datos abiertos del Perú, 2020

Hipótesis Alterna (HA):

Las redes neuronales recurrentes ayudan a predecir el aumento de la tasa de mortalidad de Covid-19 usando los datos abiertos del Perú, 2020

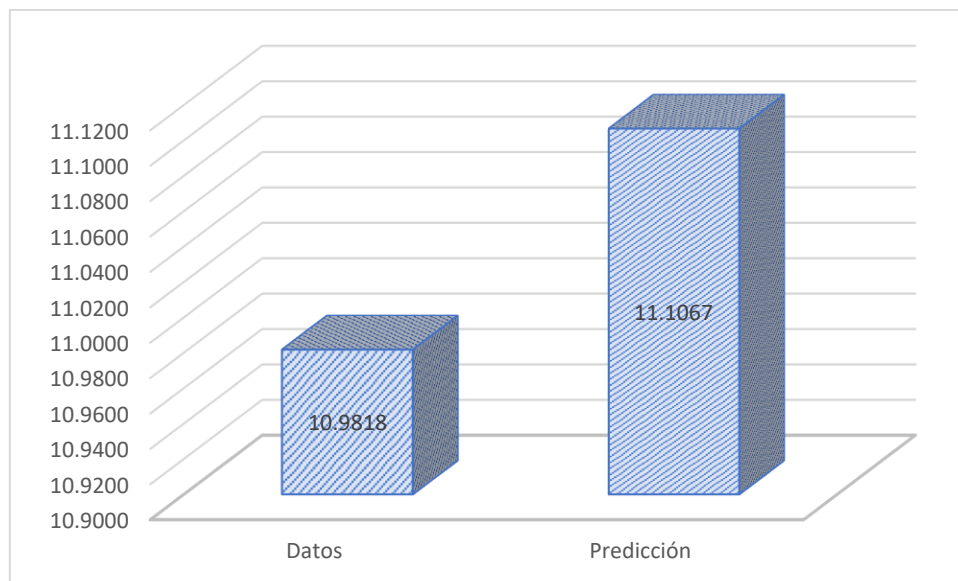
PRUEBA DESCRIPTIVA

Tabla 2: Estadísticos descriptivos de la tasa de letalidad de los datos y de la predicción de Covid-19

Estadísticos descriptivos					
	N	Mínimo	Máximo	Media	Desv. Desviación
Tasa_mortalidad	7	10,9376	11,0237	10,981800	,0314741
Tasa_mortalidad_predicción	7	11,0400	11,1810	11,106657	,0510513
N válido (por lista)	7				

En la tasa de mortalidad de los datos recogidos se obtuvo una media de 10,9818 y para los datos de la predicción 11,1066, de acuerdo a la Figura 4; estos resultados nos ayudan a observar el aumento que se da de la tasa de mortalidad de los datos y de la predicción de 7 días.

Figura 4: Grafico de barras de la Tasa de mortalidad de los datos y predicción de Covid-19



PRUEBA ESTADÍSTICA DE NORMALIDAD

Se realizó la prueba de Shapiro-Wilk, ya que se utilizó una muestra de 7 registros, lo cual es menor a 50. La muestra fue procesada por el sistema de SPSS 25, con un nivel de confiabilidad del 95% de acuerdo a las siguientes condiciones:

Si: Sig. < 0.05 es una distribución no normal
 Sig. ≥ 0.05 es una distribución normal

Tabla 3: Prueba de normalidad de la tasa de mortalidad de los datos y de la predicción de Covid-19

Pruebas de normalidad			
	Shapiro-Wilk		
	Estadístico	gl	Sig.
Tasa_mortalidad	,970	7	,898
Tasa_mortalidad_predicción	,974	7	,924

En la Tabla se visualizan los resultados de la prueba de normalidad el cual indica que el Sig. de la tasa de mortalidad de los datos de Covid-19 fue 0.898, lo cual es mayor que 0.05, por lo tanto, se confirma que tiene una distribución normal.

En la tasa de mortalidad de la predicción del Covid-19 indica que el Sig. fue 0,924, el cual es mayor que 0,05, por lo que también tiene una distribución normal. En la Figura 5 y 6 se podrá observar los histogramas de la prueba de normalidad para la tasa de mortalidad de los datos como también de la predicción del Covid-19.

Figura 5: Histograma de la tasa de mortalidad usando los datos del Covid-19

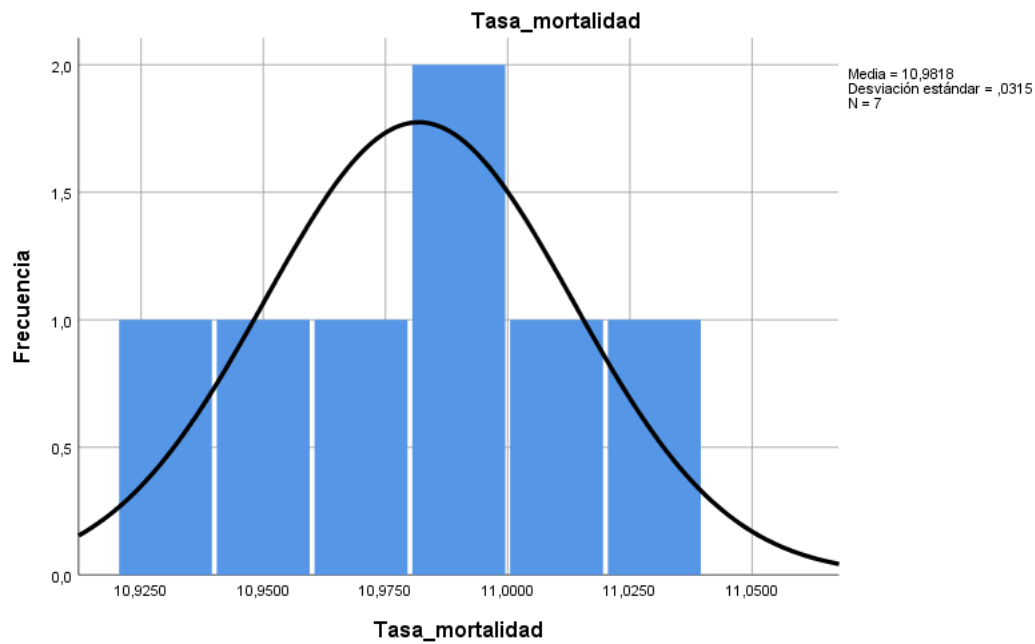
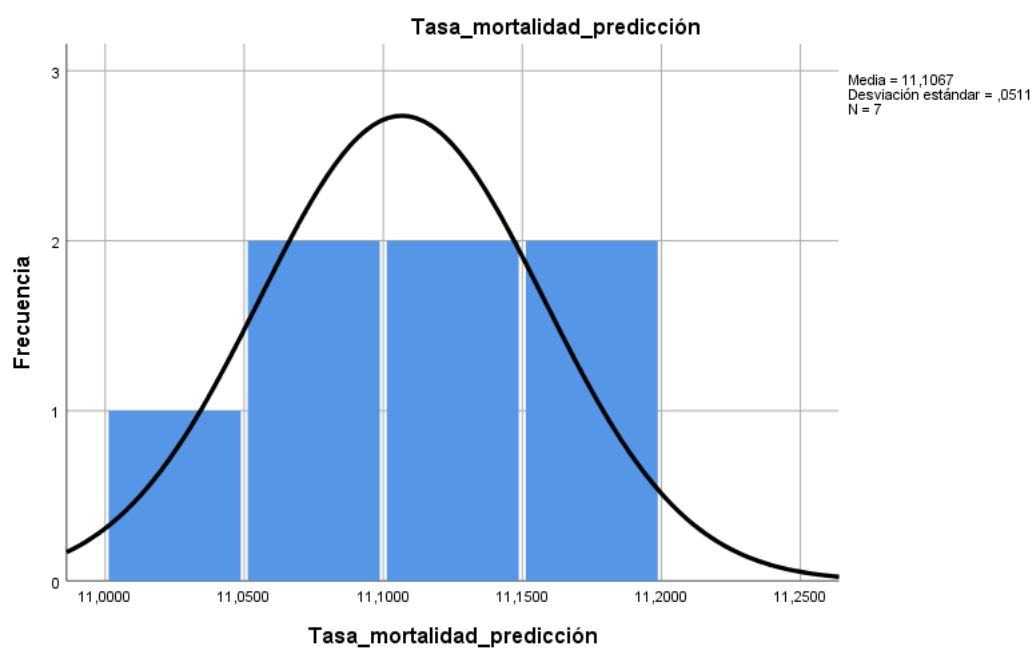


Figura 6: Histograma de la tasa de mortalidad de la predicción del Covid-19



En cuanto al resultado del contraste de hipótesis se aplicó la Prueba T-Student (Tabla 3), debido a que los datos obtenidos durante la investigación tuvieron una distribución normal. En el cual el valor de T contraste es -16,583, siendo menor que -1.9432, valor que se obtuvo de la Tabla T-Student con un grado de libertad de 6 y nivel de confianza de 0.05. Se puede verificar en el Anexo 5.

Tabla 4: Prueba T-Student para la tasa de mortalidad de los datos y predicción de Covid-19

Prueba de T-Student					
	Media	T	gl	Sig. (bilateral)	Desviación
Tasa_mortalidad	10,981,800	-16,583	6	,000	,0199206
Tasa_mortalidad_predicción	11,106,657				

Entonces, se rechaza la hipótesis nula, aceptando la hipótesis alterna con un 95% de confianza. Además, el valor T obtenido, como se muestra en la Figura 7, se ubica en la zona de rechazo. En consecuencia, Las redes neuronales recurrentes ayudan a predecir el aumento de la tasa de mortalidad de Covid-19 usando los datos abiertos del Perú, 2020.

Figura 7: Prueba de T-Student de la tasa de mortalidad



Indicador 2: Variación de la tasa de letalidad

HIPÓTESIS ESPECÍFICA

Las redes neuronales recurrentes ayudan a predecir la variación de la tasa de letalidad de Covid-19 usando los datos abiertos del Perú, 2020

HIPÓTESIS ESTADÍSTICAS

Hipótesis Nula (H0):

Las redes neuronales recurrentes no ayudan a predecir la variación de la tasa de letalidad de Covid-19 usando los datos abiertos del Perú, 2020

Hipótesis Alterna (HA):

Las redes neuronales recurrentes ayudan a predecir la variación de la tasa de letalidad de Covid-19 usando los datos abiertos del Perú, 2020

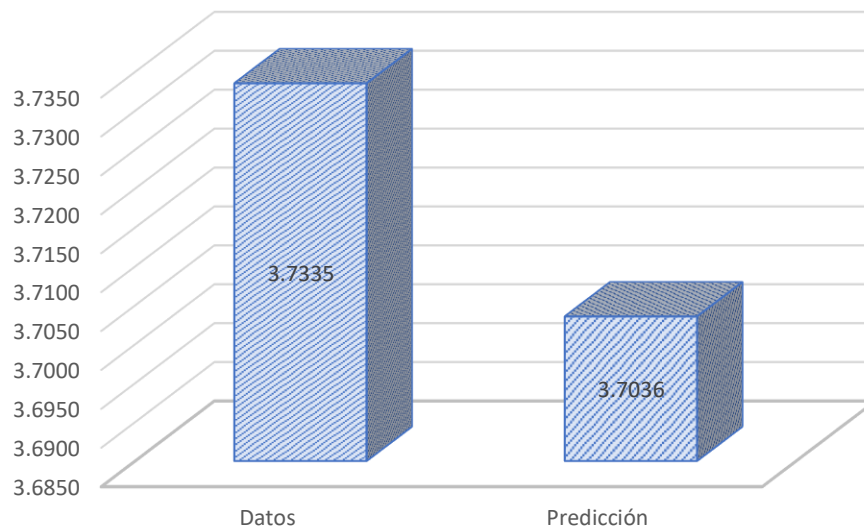
PRUEBA DESCRIPTIVA

Tabla 5: Estadística descriptiva de la tasa de letalidad de los datos y de la predicción de Covid-19

Estadísticos descriptivos					Desv. Desvia-
	N	Mínimo	Máximo	Media	ción
Tasa_letalidad	7	3,7313	3,7348	3,733529	,0012051
Tasa_letalidad_predicción	7	3,7027	3,7045	3,703571	,0006800
N válido (por lista)	7				

En la tasa de mortalidad de los datos recogidos se obtuvo una media de 3,7335 y para los datos de la predicción 3,7036, de acuerdo a la Figura 8; estos resultados nos ayudan a observar el aumento que se da de la tasa de mortalidad de los datos y de la predicción de 7 días.

Figura 8: Gráfico de barras de la Tasa de mortalidad de los datos y predicción de Covid-19



PRUEBA ESTADÍSTICA DE NORMALIDAD

Se realizó la prueba de Shapiro-Wilk, ya que se utilizó una muestra de 7 registros, lo cual es menor a 50. La muestra fue procesada por el sistema de SPSS 25, con un nivel de confiabilidad del 95% de acuerdo a las siguientes condiciones:

Si: Sig. < 0.05 es una distribución no normal

Sig. ≥ 0.05 es una distribución normal

Tabla 6: Prueba de normalidad de la tasa de letalidad de los datos y de la predicción de Covid-19

Pruebas de normalidad			
	Shapiro-Wilk		
	Estadístico	gl	Sig.
Tasa_letalidad	,908	7	,383
Tasa_letalidad_predicción	,921	7	,474

En la Tabla se visualizan los resultados de la prueba de normalidad el cual indica que el Sig. de la tasa de letalidad de los datos de Covid-19 fue 0.383, lo cual es mayor que 0.05, por lo tanto, se confirma que tiene una distribución normal. En la tasa de letalidad de la predicción del Covid-19 indica que el Sig. fue 0,474, el

cual es mayor que 0,05, por lo que también tiene una distribución normal. En la Figura 5 y 6 se podrá observar los histogramas de la prueba de normalidad para la tasa de letalidad de los datos como también de la predicción del Covid-19.

Figura 9: Histograma de la tasa de letalidad usando los datos del Covid-19

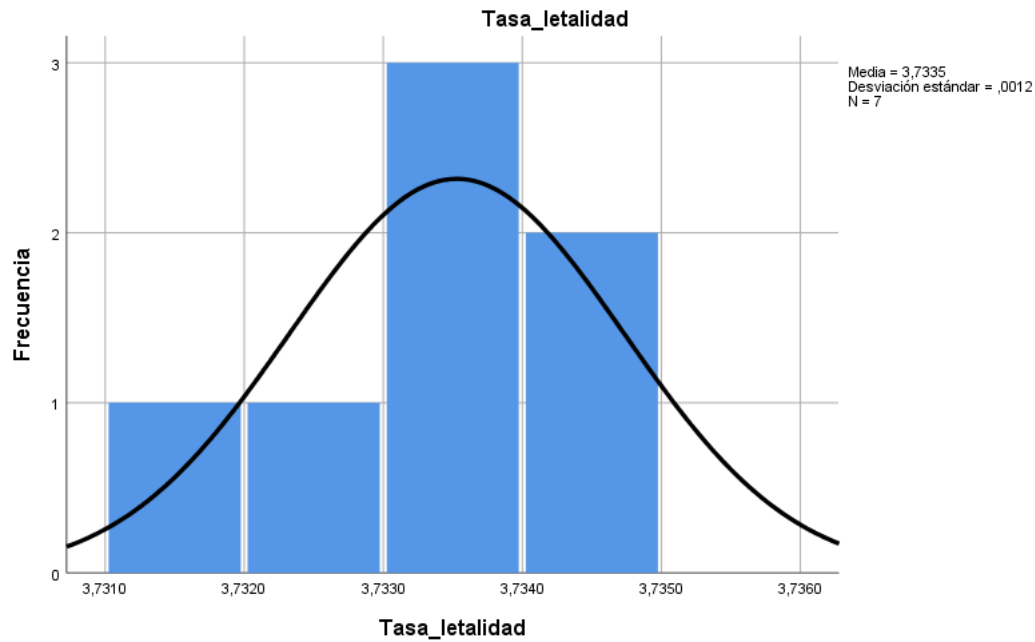
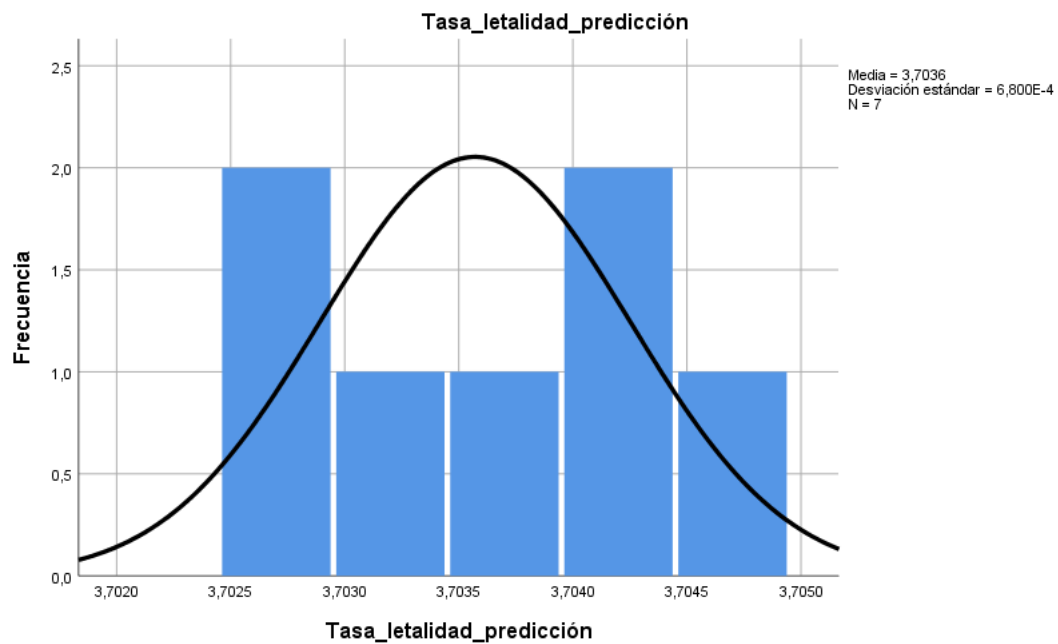


Figura 10: Histograma de la tasa de letalidad usando los datos del Covid-19



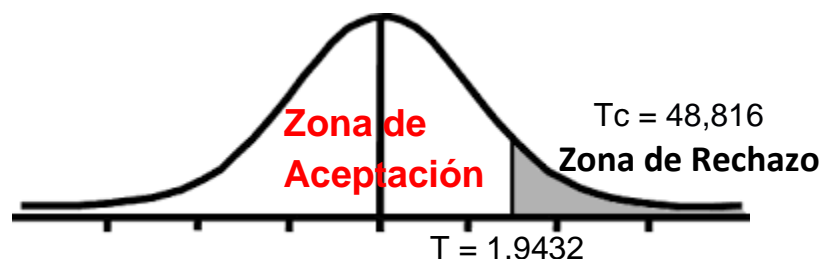
En cuanto al resultado del contraste de hipótesis se aplicó la Prueba T-Student (Tabla 3), debido a que los datos obtenidos durante la investigación tuvieron una distribución normal. En el cual el valor de T contraste es 48,816, siendo mayor que 1.9432, valor que se obtuvo de la Tabla T-Student con un grado de libertad de 6 y nivel de confianza de 0.05. Se puede verificar en el Anexo 5.

Tabla 7: Prueba T-Student para la tasa de letalidad de los datos y predicción de Covid-19

Prueba de T-Student					
	Media	T	gl	Sig. (bilateral)	Desviación
Tasa_letalidad	3,7335	48,816	6	,000	,0016236
Tasa_letalidad_predicción	3,7036				

Entonces, se rechaza la hipótesis nula, aceptando la hipótesis alterna con un 95% de confianza. Además, el valor T obtenido, como se muestra en la Figura 7, se ubica en la zona de rechazo. En consecuencia, Las redes neuronales recurrentes ayudan a predecir el aumento de la tasa de letalidad de Covid-19 usando los datos abiertos del Perú, 2020.

Figura 11: Prueba de T-Student de la tasa de letalidad



V. DISCUSIÓN

A continuación, se describen los resultados de la presente investigación en el cual se ha basado de la predicción de casos de Covid-19 de 7 días utilizando los datos de la plataforma de datos abiertos de Perú para calcular el aumento de la tasa de mortalidad y la variación de la tasa de letalidad. En nuestra predicción se utilizaron los datos de las fechas del 23 al 29 de noviembre del 2020 como datos de entrada para nuestro modelo predictivo, el cual da como resultado la predicción de los casos de los días del 30 de noviembre al 6 de diciembre del 2020.

En nuestros resultados se observa que el aumento medio de la tasa de mortalidad de Covid-19 por cada 10 000 habitantes entre los datos recogidos (10.9818) y la predicción (11,1057), lo cual muestra que el aumento de la tasa de mortalidad ha sido de 0,1248 por cada 10 000 habitantes.

Luego están los resultados de la variación de la tasa de letalidad de Covid-19 entre los datos recogidos (3,7335) y la predicción (3,7036), lo cual indica que hubo una variación media del 0.03 % en la tasa de letalidad.

Para optimizar el modelo predictivo de los casos por Covid-19 se utilizó el método para que utilizaron los autores (Flores, Tito y Cently, 2020) en el cual realiza una tabla para definir el tipo de arquitectura que predice con el RMSE más bajo, dando como resultado que en la mayoría de los casos las arquitecturas híbridas son las que dan un RMSE menor a diferencia de las no híbridas. En nuestra investigación el RMSE más bajo se obtuvo de la arquitectura híbrida (BI-GRU – BI-LSTM). Por otra parte, los autores (Shastri, Singh, Kumar, Kour y Mansotra) desarrollaron una predicción de casos confirmados y fallecidos por Covid-19 de India y EEUU, en el cual utilizan el error absoluto medio porcentual (MAPE) para medir el error de su predicción, en el cual los casos fallecidos en India tienen un mayor error (4,00) con respecto a la predicción de los confirmados (3,33) y en EEUU ocurre lo contrario; la predicción de los casos confirmados obtuvo un mayor error (6,66) con respecto a los casos fallecidos (6,00). En nuestra investigación el error se calcula el error mediante el RMSE con una escala de 0 a 1, del cual se obtuvo que para los casos fallecidos tuvieron un error (0,086) más alto con respecto a los casos de confirmados (0,045). Para validar nuestra investigación, se realizó el estado del arte en el cual se ha registrado más de 50

investigaciones de artículos indexados y tesis en las cuales no se encontraron investigaciones que realicen predicciones de los casos de Covid-19 en el Perú utilizando las redes neuronales recurrentes, por lo que no hay forma de comparar el resultado que se obtuvo del RMSE, por lo que nos basamos en el autor (Zevallos, 2017) quien mencionó que “El nivel de predicción es de $PRED(0.16) = 0.84$. Sabiendo que para que un modelo predictivo sea aceptable un PRED debe ser mayor a 0.75”. Por lo que si restamos 1 – menos el valor mencionado (0.75) obtenemos el margen de error de una predicción aceptable se encuentra dentro de 0,25, el error que se obtuvo del RMSE para los casos confirmados y fallecidos fueron de 0.0445 y 0,0861 respectivamente, estos resultados están en una escala de 0 a 1 por lo que en porcentaje se representaría 4.45 % para los casos confirmados y 8.61 % para los casos fallecidos, mostrando así que nuestra predicción está dentro de un margen de error aceptable.

VI. CONCLUSIONES

Las conclusiones de la investigación fueron las siguientes:

1. Se concluye que la hipótesis alternativa Las redes neuronales recurrentes ayudan a predecir el aumento de la tasa de mortalidad de Covid-19 usando los datos abiertos del Perú, 2020 es aceptada con un 95% de confianza al realizar la prueba T de Student el cual el valor de T contraste es -16,583, siendo menor que -1.9432, valor que se obtuvo de la Tabla T-Student con un grado de libertad de 6 y nivel de confianza de 0.05.
2. Se concluye que la hipótesis alternativa Las redes neuronales recurrentes ayudan a predecir la variación de la tasa de letalidad de Covid-19 usando los datos abiertos del Perú, 2020 es aceptada con un 95% de confianza al realizar el T de Student el cual el valor de T contraste es 48,816, siendo mayor que 1.9432, valor que se obtuvo de la Tabla T-Student con un grado de libertad de 6 y nivel de confianza de 0.05.
3. El resultado del indicador tasa de mortalidad según la prueba de normalidad de acuerdo al método shapiro-wilk, los datos tienen una distribución normal ya que la muestra indica que el Sig. fue 0,924 como indica dicho método.
4. El resultado del indicador tasa de letalidad según la prueba de normalidad de acuerdo al método shapiro-wilk, los datos tienen una distribución normal ya que la muestra indica que el Sig. fue 0,474 como indica dicho método.
5. Del objetivo de utilizar las redes neuronales recurrentes para predecir el aumento de la tasa de mortalidad de casos por Covid-19 usando datos abiertos en el Perú, 2020, se utilizaron las redes neuronales para predecir el aumento medio de la tasa de mortalidad en 0,1248 por cada 10 000 habitantes.
6. Del objetivo utilizar las redes neuronales recurrentes para predecir la variación de la tasa de letalidad de casos por Covid-19 usando datos abiertos en el Perú, 2020, se utilizaron las redes neuronales para predecir la

variación media de la tasa de letalidad, en el cual se obtuvo como resultado del 0,03%.

7. Del objetivo utilizar las redes neuronales recurrentes para predecir los casos por Covid-19 usando datos abiertos en el Perú, 2020, se utilizaron las redes neuronales para predecir el aumento de los casos de Covid-19. Al utilizar las RNN híbridas permitieron obtener una predicción de casos confirmados como de fallecidos que a pesar de tener una cantidad limitada de entradas de registros para su entrenamiento proporciona cifras que se acercan a la realidad para sacar la tasa de letalidad y mortalidad.

VII. RECOMENDACIONES

Las recomendaciones para futuras investigaciones son las siguientes:

- 1 Las redes neuronales recurrentes son de aprendizaje profundo lo cual requiere grandes cantidades de datos para que logre un buen aprendizaje por ende disminuya el porcentaje de error de manera que la red neuronal pueda entrenar más. Se recomienda recolectar más datos para que la predicción sea más confiable y eficaz.
- 2 Para realizar una predicción de series de tiempo con los datos de confirmados y fallecidos por Covid-19 que se encuentran en la plataforma de datos abiertos del Perú, se recomienda aplicar distintas técnicas para el tratamiento de datos.
- 3 Para utilizar las redes neuronales recurrentes se debe tener en cuenta la cantidad de datos que se tiene disponible y la cantidad de épocas con las que se va entrar el modelo.
- 4 Trabajar con las redes neuronales recurrentes ha sido un tema muy interesante con mucho por investigar y aprender para lograr tener una buena predicción, por lo que se motiva a los investigadores a seguir realizando este tipo de investigaciones.
- 5 Se recomienda que se amplíe la investigación para que en un futuro sirva de base de conocimientos y permita el discernimiento de resultados gracias al nivel de confiabilidad de las predicciones.

REFERENCIAS

A

AGARWAL, Basant. *Deep Learning-Based Approaches for Sentiment Analysis*. Springer Nature, 2020. ISBN 9789811512162

AL-NAJJAR, H. y AL-ROUSAN, N., 2020. A classifier prediction model to predict the status of Coronavirus CoVID-19 patients in South Korea. *European Review for Medical and Pharmacological Sciences*, vol. 24, no. 6, pp. 3400-3403. ISSN 22840729. DOI 10.26355/eurrev_202003_20709.

ALTAMIRANO Pontigo, SERGIO.2020. Esteban. Aplicación de redes neuronales recurrentes y modelos de series de tiempo bayesianos a la predicción de rentabilidad de fondos de pensiones. Universidad de Concepción, Facultad de Ciencias Física y Matemáticas, Departamento de Estadística. Tesis para optar al título profesional de Ingeniero. Disponible en: <http://repositorio.udec.cl/xmlui/handle/11594/526>

ÁLVARES, Rafael. *El método científico en las ciencias de la salud*. España. Ediciones Díaz de Santos, 1996. ISBN: 9788479782375

Alocución del Director General de la OMS en la rueda de prensa sobre la COVID-19 celebrada el 24 de febrero de 2020. Dr Tedros Adhanom Ghebreyesus 23 de febrero de 2020. Disponible en : <https://www.who.int/es/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---24-february-2020>

APESTEGUIA, Juan y HUARCAYA, Edwin. Redes neuronales artificiales aplicadas a la detección de fugas no visibles en las redes de agua potable de la ciudad de Lima. Tesis para optar el grado de Maestro en Ciencias de la Electrónica. Lima: Universidad del Callao, 2018. Disponible en: <http://repositorio.unac.edu.pe/handle/UNAC/3899>

B

BOTTO, Miguel, LEÓN, Joffre, DÍAS, Angela y MONTIEL, Práxedes. *Advances in Emerging Trends and Technologies: Volume 1*. Springer Nature, 2020. ISBN: 9783030320225

BONACCORSO, Giuseppe. *Hands-on Unsupervised Learning with Python*. Birmingham. Published by Packt Publishing Ltd, 2019. ISBN 978-1-78934-827-9

BERNAL, César Augusto. Metodología de la investigación: para administración, economía, humanidades y ciencias sociales, Mexico, Pearson Educación, 2006. ISBN: 9789702606451.

C

CAÑADAS Isabel y SAN LUIS Concepción. *Análisis de datos en investigación. Primeros pasos*. España. Universidad Miguel Hernández, 2018. ISBN: 9788416024742

Center for Systems and Engineering (CSSE). Publicado el 24 de abril del 2020. [Fecha de consulta: 24 de abril]. Disponible en: <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>

CIABURRO, Giuseppe y VENTESWARAN, Balaji. *Neural Networks with R*. Birmingham. Packt Publishing Ltd, 2017. ISBN 9781788397872

CHAKRAVERTY, Snehashish y MALL, Susmita. *Artificial Neural Networks for Engineers and Scientists: Solving Ordinary Differential Equations*. CRC Press, 2017. ISBN: 9781498781404

CÓRDOVA, Edwin y MAUTINO, Jhojan. Análisis de redes neuronales artificiales para el pronóstico de las enfermedades prevalentes en el hospital Víctor Ramos Guardia de Huaraz, 2016. Tesis para optar al grado de Licenciado en Estadística e Informática. Lima: Universidad Nacional Santiago Antúnez de Mayolo, 2019. Disponible: <http://repositorio.unasam.edu.pe/handle/UNASAM/3545>

CORONAVIRUS: cómo Estados Unidos se convirtió en el nuevo centro de la pandemia de covid-19. Recursos internet [en línea]. BBC News Mundo. Publicado el 1 de abril del 2020. [Fecha de consulta: 22 de abril]. Disponible en: <https://www.bbc.com/mundo/noticias-internacional-52114455>

CRONOLOGÍA del coronavirus: así comenzó y se extendió el virus que tiene en alerta al mundo. Recursos internet [en línea]. CNN Español. Publicado el 20 de febrero 2020. [Fecha de consulta: 22 de abril]. Disponible en: <https://cnnespanol.cnn.com/2020/02/20/cronologia-del-coronavirus-asi-comenzo-y-se-extendio-el-virus-que-pone-en-alerta-al-mundo/>

CORONAVIRUS COVID-19 (2019-nCoV). Recursos internet [en línea].

CUEVA, Percy. Evaluación de índices de confiabilidad en redes de Electrocentro S.A, para mejorar la calidad de servicio utilizando redes neuronales. Tesis para optar para doctor en Ingeniería eléctrica. Callao: Universidad Nacional del Callao, 2019. Recuperado en: <http://repositorio.unac.edu.pe/handle/UNAC/4474>

D

DING, W., XIA, Y., WANG, Z., CHEN, Z. y GAO, X., 2020. An ensemble-learning method for potential traffic hotspots detection on heterogeneous spatio-temporal data in highway domain. *Journal of Cloud Computing*, vol. 9, no. 1, pp. 1-11. ISSN 2192113X. DOI 10.1186/s13677-020-00170-1.

E

EDGAR SERNA M. *Desarrollo e Innovación en Ingeniería*. 2da Edición. Medellín: Editorial Instituto Antioqueño de Investigación, 2017. ISBN: 978-958-59127-5-5

EL-SHAHAT, Adel. *Advanced Applications for Artificial Neural Networks*. Rijeka: IntechOpen, 2018. ISBN: 9789535137801

El estudio y la investigación documental: Estrategias metodológicas y herramientas TIC por Simona María Parraguez Carrasco [et al.]. Chiclayo: Biblioteca Nacional del Perú, 2017. ISBN: 9786120026038

ESCOBAR, Piter y BILBAO Jorge. *Investigación y educación superior*. Segunda Edición. EE. UU: Lulu.com, 2020. ISBN: 9781678103903

F

FANELLI, D. y PIAZZA, F., 2020. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons and Fractals*, vol. 134. ISSN 09600779. DOI 10.1016/j.chaos.2020.109761.

FLORES A., TITO H., Centty D. 2020. Comparison of Hybrid Recurrent Neural Networks for Univariate Time Series Forecasting. In: Arai K., Kapoor S., Bhatia R. (eds) *Intelligent Systems and Applications. IntelliSys 2020. Advances in Intelligent Systems and Computing*, vol 1250. Springer, Cham. http://doi-org-443.webvpn.fjmu.edu.cn/10.1007/978-3-030-55180-3_28

G

GRANDE y ABASCAL. *Fundamentos y técnicas de investigación comercial*. Madrid: ESIC Editorial, 2011. 435 pp. ISBN: 9788473567473

GRÁNDEZ, Miguel. Aplicación de minería de datos para determinar patrones de consumo futuro en clientes de una distribuidora de suplementos nutricionales. Tesis para optar para Título Profesional de Ingeniero Informático y de sistemas. Universidad San Ignacio de Loyola. Perú. Recuperado: <http://repositorio.usil.edu.pe/handle/USIL/2763>.

H

HERNANDEZ, Roberto, FERNÁNDEZ, Carlos y BAPTISTA, María. *Metodología de la investigación*. 6ta Edición. México: McGraw-Hill Interamericana Editores, S.A. de C.V., 2014. ISBN: 9781456223960

HERNÁNDEZ, Roberto y MENDOZA, Christian. *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. México: McGraw-Hill Interamericana Editores, S.A. de C.V., 2018. ISBN: 9781456260965

HEN HU, Yu. *Handbook of Neural Network Signal Processing*. EE.UU. Taylor & Francis Inc CRC Press, 2001. ISBN: 9780849323591

HERNÁNDEZ, Benjamín. *Técnicas estadísticas de investigación social*. Madrid: Ediciones Díaz de Santos, 2001. ISBN: 9788479785055

He, W., Yi, G. Y., & Zhu, Y. (2020). Estimation of the basic reproduction number, average incubation time, asymptomatic infection rate, and case fatality rate for COVID-19: Meta-analysis and sensitivity analysis. *Journal of Medical Virology*. doi:10.1002/jmv.26041

HUNDT Marianne, MOLLIN Sandra y E. PFENNIGER Simone. *The Changing English Language: Psycholinguistic Perspectives*. EE.UU. Cambridge University Press, 2017. ISBN: 9781107086869

J

JOSHI, Prateek. *Artificial Intelligence with Python*. Birmingham. Published by Packt Publishing Ltd, 2017. ISBN: 9781786469670

K

KOSTADINOV, Simeon. *Recurrent Neural Networks with Python Quick Start Guide*. Birmingham. Published by Packt Publishing Ltd, 2018. ISBN 978-1-78973-233-5

L

LONGEST, Kyle. *Using Stata for Quantitative Analysis*. 3ra Edición. Londres: SAGE Publications, 2019. ISBN: 9781544318530

M

MOFTAKHAR, L., SEIF, M. y SAFE, M.S., 2020. Exponentially increasing trend of infected patients with covid-19 in iran: A comparison of neural network and arima forecasting models. *Iranian Journal of Public Health*, vol. 49, pp. 92-100. ISSN 22516093.

MORENO, A., LÓPEZ, S. y CORCHO, A., 2015. Principales medidas en epidemiología. *Salud Pública de México* [en línea], vol. 42, no. 4, pp. 338-348. Disponible en: https://www.scielo.org/article/ssm/content/raw/?resource_ssm_path=/media/assets/spm/v42n4/2882.pdf.

MOTA, Arantxa, 2016. Pronóstico del Precio del Crudo de Exportación Mexicano con la Metodología de Box Jenkins para Series de Tiempo. Tesis para optar título de Licenciada en Actuaría, Benemérita Universidad Autónoma de Puebla Disponible en: <https://1library.co/document/eqp2o9ky-pronostico-precio-exportacion-mexicano-metodologia-jenkins-series-tiempo.html#reference-content>.

N

NAGALES, Ángel. *Investigación y técnicas de mercado*. 2.a ed. Madrid: ESIC Editorial, 2004. ISBN: 9788473563925

NEIRA, Mike. 2017, Descripción metodológica de las series de tiempo con redes neuronales artificiales. Tesis para optar el título de Ingeniero Estadístico e

Informático, Universidad Nacional Agraria La Molina. Disponible en: <http://repositorio.lamolina.edu.pe/handle/UNALM/3276?show=full>

NORMAN, Alan. *Aprendizaje automático en acción*. Editorial Tedtime, 2017. ISBN: 9788893988742

Ñ

ÑAUPAS, H., VALDIVIA, M., PALACIOS, J. y ROMERO, H. *Metodología de la Investigación cuantitativa-cualitativa y redacción de la tesis*. 5ta Edición. Bogotá: Ediciones de la U, 2018. ISBN: 9789587628777

P

PATAN, Krzysztof. *Robust and Fault-Tolerant Control: Neural-Network-Based Solutions*. Góra: Springer, 2019. ISBN: 9783030118693

PÉREZ Elena, LIBERATORE Federico. 2018. "Predicción demanda eléctrica española. Implementación de redes neuronales recurrentes en Python. Universidad Complutense de Madrid España. Tesis para Trabajo de fin de máster en Ingeniería Matemática. Disponible en: <https://eprints.ucm.es/49444/>

PÉREZ, Rigoberto y LÓPEZ, Ana. *Métodos estadísticos para Economía y Empresa*. España: Universidad de Oviedo, 2011. ISBN: 9788469490099

PIROUZ, B., HAGHSHENAS, Sina Shaffiee, HAGHSHENAS, Sami Shaffiee y PIRO, P., 2020. Investigating a serious challenge in the sustainable development process: Analysis of confirmed cases of COVID-19 (new type of Coronavirus) through a binary classification using artificial intelligence and regression analysis. *Sustainability (United States)*, vol. 12, no. 6. ISSN 19370695. DOI 10.3390/su12062427.

PUENTE Jorge, De OLIVEIRA Celma, RUELAS Edgar. 2020. Pronóstico de la demanda de combustible de etanol en Brasil uso de una red neuronal recurrente LSTM. *IEEE Latin America Transactions*. Disponible en: <https://lat-amt.ieeeer9.org/index.php/transactions/article/view/3713>

PURKAIT, Niloy. *Hands-On Neural Networks with Keras: Design and create neural networks using deep learning and artificial intelligence principles*. Birmingham. Packt Publishing, 2019. ISBN: 978-1-78953-608-9

R

RAJENDRA Akerkar, PRITI Sajja. *Knowledge-Based Systems*. Sudbury. Massachusetts Jones & Bartlett Learning, 2010. ISBN: 9780763776473

RODRÍGUEZ, Noelia. Pronóstico del caudal máximo del Río Huaura por modelos Arima y redes neuronales. Tesis para optar el título profesional de Ingeniero Estadístico. Trujillo: Universidad Nacional de Trujillo, 2017. Recuperado en: <http://dspace.unitru.edu.pe/handle/UNITRU/10637>

Ruiz Lopez, H. R. (2018). Pronóstico de las exportaciones del cacao ecuatoriano para el 2018 con el uso de modelos de series de tiempo. *INNOVA Research Journal*, 3(6), 9-20. <https://doi.org/10.33890/innova.v3.n6.2018.713>

S

SANCHEZ, H. , Reyes C. y Mejía K. Manual de términos en investigación científica, tecnológá y humanística. Perú: ©Universidad Ricardo Palma, 2018. ISBN: 978-612-47351-4-1

SCHWARZ, M., 2017. Redes Neuronales para el reconocimiento de patrones no tradicionales en la predicción de accidentes fatales en la industria minera peruana. *Revista Científica de la UCSA*, 4(2), 6-12. Recuperado de: [http://ucsa.edu.py/yeah/wp-content/uploads/2017/09/2.-AO Schwarz-M 6-12-GV-sin-doi.pdf](http://ucsa.edu.py/yeah/wp-content/uploads/2017/09/2.-AO_Schwarz-M_6-12-GV-sin-doi.pdf)

Shastri, S., Singh, K., Kumar, S., Kour, P., & Mansotra, V. (2020). Time Series Forecasting of Covid-19 using Deep Learning Models: India-USA Comparative Case Study. *Chaos, Solitons & Fractals*, 110227. doi:10.1016/j.chaos.2020.110227

SINALUISA, Ivan. Aplicación de La Red Neuronal Artificial Feedforward Back-propagation para la predicción de demanda de energía eléctrica en la Empresa Eléctrica Riobamba S.A. Tesis para optar Magíster en Sistemas de control y Automatización Industrial. Chimborazo: Escuela Superior Politécnica de Chimborazo, 2017. Recuperado en: <http://dspace.es-poch.edu.ec/handle/123456789/7606>

SKANSI, Sandro. *Introduction to Deep Learning: From Logical Calculus to Artificial Intelligence*. Switzerland. Springer, 2018. ISBN: 9783319730042

SOHRABI, C., ALSAFI, Z., O'NEILL, N., KHAN, M., KERWAN, A., AL-JABIR, A., IOSIFIDIS, C. y AGHA, R., 2020. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). 1 abril 2020. S.l.: Elsevier Ltd.

SOLÍZ, Desiderio. *Cómo Hacer Un Perfil Proyecto De Investigación Científica*. Palibrio, 2019. ISBN 9781506527208

T

TA, Van-Dai, LIU, CHUAN-MING and TADESSE, Direselign Addis. Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading. *Applied Sciences* [online]. 7 January 2020. Vol. 10, no. 2, p. 437. DOI 10.3390/app10020437. Available from: <http://dx.doi.org/10.3390/app10020437>

TRASK, Andrew. *Grokking Deep Learning*. Manning Publications Co., 2019. ISBN: 9781617293702

Tölö, E. (2020). Predicting systemic financial crises with recurrent neural networks. *Journal of Financial Stability*, 100746. doi:10.1016/j.jfs.2020.100746

TORO, Iván y PARRA, Rubén. *Método y conocimiento: metodología de la investigación: investigación cualitativa/investigación cuantitativa*. Medellín: Universidad Eafit, 2006. ISBN: 9789588281117

TORRES, Jordi. *DEEP LEARNING Introducción práctica con Keras*. 3ra Edición. Universidad Politécnica de Catalunya, 2018. ISBN: 9780244078959

U

Un nuevo coronavirus, una nueva enfermedad: COVID-19 por Quiroz [et al.]. *Horizonte Médico* (Lima), [S.l.], v. 20, n. 2, p. e1208, abr. 2020. ISSN 2227-3530.

W

WANG, H., WANG, Z., DONG, Y., CHANG, R., XU, C., YU, X., ZHANG, S., TSAMLAG, L., SHANG, M., HUANG, J., WANG, Y., XU, G., SHEN, T., ZHANG, X. y CAI, Y., 2020. Phase-adjusted estimation of the number of Coronavirus Disease 2019 cases in Wuhan, China. *Cell Discovery*, vol. 6, no. 1. ISSN 20565968. DOI 10.1038/s41421-020-0148-0.

WELLS, Andrew. *Rethinking Cognitive Computation: Turing and the Science of the Mind*. EE.UU. Red Globe Press, 2017. ISBN: 9781137066619.

Z

Zambrano, O., Espitia, S., & Hernández, J. (2017). Cultura del emprendimiento en instituciones de educación superior: estrategia de inclusión social. *Desarrollo Gerencial*, 9(1), 176-191. <https://doi.org/10.17081/dege.9.1.2732>

ZEBALLOS, Rodolfo. Predicción del rendimiento académico mediante redes neuronales. Tesis para optar grado Ingeniero Electrónico. Callao: Universidad Nacional del Callao, 2017. Disponible: <http://repositorio.unac.edu.pe/handle/UNAC/2728>

ZHAO, Z., LI, X., LIU, F., ZHU, G., MA, C. y WANG, L., 2020. Prediction of the COVID-19 spread in African countries and implications for prevention and controls: A case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya. *Science of The Total Environment* [en línea], vol. 729, pp. 138959. ISSN 00489697. DOI 10.1016/j.scitotenv.2020.138959. Disponible en: <https://doi.org/10.1016/j.scitotenv.2020.138959>.

ANEXO 1: MATRIZ DE OPERACIONALIZACIÓN DE VARIABLES

Tabla 8: Matriz de operacionalización de variables

Variable Dependiente	Definición conceptual	Definición operacional	Dimensión	Indicadores	Instrumento	Escala de Medición
Redes neuronales recurrentes	En particular, las redes neuronales recurrentes son excelentes para modelar datos secuenciales. [...] Cuando trabajamos con datos de series de tiempo, no podemos simplemente usar modelos de aprendizaje genéricos. Necesitamos caracterizar las dependencias temporales en nuestros datos para que podamos construir un modelo robusto. (Joshi, 2017)	La variable se utilizará de apoyo para obtener los resultados de la predicción de 7 días a futuro. (Joshi, 2017).				

Variable dependiente	Definición conceptual	Definición operacional	Dimensión	Indicadores	Instrumento	Escala de Medición
Predicción de casos por Covid-19 Según Fanelli y Piazza (2020)	Según Fanelli y Piazza (2020) indica que la predicción de casos servirá para “proporcionar a los funcionarios estimaciones realistas del tiempo y la magnitud del pico epidémico, es decir, el número máximo de personas infectadas [...]. Los datos se obtuvieron del repositorio de GitHub asociado con el tablero interactivo alojado por el Centro de Ciencia e Ingeniería de Sistemas (CSSE) en la Universidad Johns Hopkins” (p. 1).	La variable se analiza en función a dos indicadores los cuales son la tasa de mortalidad del cual es conocer el cociente entre el número total de muertes ocurridas en un país en un periodo determinado y el número de habitantes de la población de referencia y la tasa de letalidad para conocer el número de personas fallecidas por una determinada enfermedad en un periodo de tiempo determinado. Fanelli, D. y Piazza 2020.	Predicción de casos (Mohtakhar, Seif y Safe, 2020)	Tasa de mortalidad x 100000 (TM): $TM = \frac{TF}{NH} \times 100000$ TF = Total de fallecidos por Covid-19 NP = Número de habitantes en el Perú (Asencios, 2015)	Técnica: Fichaje Instrumento: Ficha de registro	Razón
				Tasa de letalidad (TL): $TL = \frac{TF}{TC} \times 100\%$ TF = Total de casos fallecidos por Covid-19 TC = Total de casos confirmados por Covid-19 (He, Yi y Zhu, 2020)		Razón

TABLA 9 MATRIZ DE CONSISTENCIA

Problema General	Objetivo General	Hipótesis General
¿Cómo utilizar redes neuronales recurrentes para la predicción de casos por Covid-19 usando datos abiertos en el Perú, 2020?	Utilizar las redes neuronales recurrentes para predecir los casos por Covid-19 usando datos abiertos en el Perú, 2020	Las redes neuronales recurrentes ayudan a predecir el aumento de los casos por Covid-19 usando los datos abiertos del Perú, 2020
Problemas Específicos	Objetivos Específicos	Hipótesis Específico
¿Cómo utilizar las redes neuronales recurrentes para predecir el aumento de la tasa de mortalidad de casos por Covid-19 usando datos abiertos en el Perú, 2020?	Utilizar las redes neuronales recurrentes para predecir el aumento de la tasa de mortalidad de casos por Covid-19 usando datos abiertos en el Perú, 2020	Las redes neuronales recurrentes ayudan a predecir el aumento de la tasa de mortalidad de Covid-19 usando los datos abiertos del Perú, 2020
¿Cómo utilizar las redes neuronales recurrentes para predecir la variación de la tasa de letalidad de casos por Covid-19 usando datos abiertos en el Perú, 2020?	Utilizar las redes neuronales recurrentes para predecir la variación de la tasa de letalidad de casos por Covid-19 usando datos abiertos en el Perú, 2020	Las redes neuronales recurrentes ayudan a predecir la variación de la tasa de letalidad de Covid-19 usando los datos abiertos del Perú, 2020

Anexo 2: Instrumento de recolección de datos

FIGURA 12 REGISTRO TASA DE MORTALIDAD DE LOS DATOS DE COVID-19



UNIVERSIDAD CÉSAR VALLEJO

Ficha de registro del indicador tasa de mortalidad

<u>Autores</u>	Catunta Esquivel Saulo y Roncal Hidalgo Dalyuska Mavel
<u>Indicador</u> $TM = \frac{TF}{NH} \times 100000$	TF = Total de fallecidos por Covid-19 NH = Número de habitantes en el Perú (Asencios, 2015)
<u>Objetivo</u>	Calcular tasa de mortalidad en el Perú en tiempos del Covid-19
<u>Periodo</u>	7 últimos registros estratificadas en días
<u>Fecha de aprobación</u>	29/11/2020

Nº	Fecha	Fallecidos acumulados	Nº de habitantes	Tasa de mortalidad
1	2020-11-23	35685	32625948	10,9376
2	2020-11-24	35727	32625948	10,9505
3	2020-11-25	35785	32625948	10,9683
4	2020-11-26	35839	32625948	10,9848
5	2020-11-27	35879	32625948	10,9971
6	2020-11-28	35923	32625948	11,0106
7	2020-11-29	35966	32625948	11,0237

FIGURA 13 TASA DE LETALIDAD DE LOS DATOS DE COVID 19



Ficha de registro del indicador tasa de letalidad

<u>Autores</u>	Catunta Esquivel Saulo y Roncal Hidalgo Dalyuska Mavel
<u>Indicador</u> $TL = \frac{TF \times 100}{TC}$	TL = Tasa de letalidad TF = Total del número de casos fallecidos TC = Total del número de casos confirmados (He, Yi y Zhu, 2020)
<u>Objetivo</u>	Calcular la tasa de letalidad del Covid-19 en porcentaje.
<u>Periodo</u>	7 últimos registros estratificadas en días
<u>Fecha de aprobación</u>	29/11/2020

Nº	Fecha	Confirmados acumulados	Fallecidos acumulados	Tasa de letalidad
1	2020-11-23	955477	35685	3.7348
2	2020-11-24	956955	35727	3.7334
3	2020-11-25	958415	35785	3.7338
4	2020-11-26	959886	35839	3.7337
5	2020-11-27	961564	35879	3.7313
6	2020-11-28	962338	35923	3.7329
7	2020-11-29	963005	35966	3.7348

FIGURA 14 TASA DE MORTALIDAD DE PREDICCIÓN DE COVID 19



UNIVERSIDAD CÉSAR VALLEJO

Ficha de registro del indicador tasa de mortalidad

<u>Autores</u>	Catunta Esquivel Saulo y Roncal Hidalgo Dalyuska Mavel
<u>Indicador</u> $TM = \frac{TF}{NH} \times 100000$	TF = Total de fallecidos por Covid-19 NH = Número de habitantes en el Perú (Asencios, 2015)
<u>Objetivo</u>	Calcular tasa de mortalidad en el Perú en tiempos del Covid-19
<u>Periodo</u>	7 últimos registros estratificadas en días
<u>Fecha de aprobación</u>	29/11/2020

<u>N°</u>	<u>Fecha</u>	<u>Fallecidos acumulados</u>	<u>N° de habitantes</u>	<u>Tasa de mortalidad</u>
1	2020-11-30	36019	32625948	11,04
2	2020-12-01	36077	32625948	11,0578
3	2020-12-02	36153	32625948	11,0811
4	2020-12-03	36232	32625948	11,1053
5	2020-12-04	36307	32625948	11,1283
6	2020-12-05	36388	32625948	11,1531
7	2020-12-06	36479	32625948	11,181

FIGURA 15 TASA DE LETALIDAD DE PREDICCIÓN DE COVID 19



Ficha de registro del indicador tasa de letalidad

<u>Autores</u>	Catunta Esquivel Saulo y Roncal Hidalgo Dalyuska Mavel
<u>Indicador</u> $TL = \frac{TF \times 100}{TC}$	TL = Tasa de letalidad TF = Total del número de casos fallecidos TC = Total del número de casos confirmados (He, Yi y Zhu, 2020)
<u>Objetivo</u>	Calcular la tasa de letalidad del Covid-19 en porcentaje.
<u>Periodo</u>	7 últimos registros estratificadas en días
<u>Fecha de aprobación</u>	29/11/2020

N°	Fecha	Confirmados acumulados	Fallecidos acumulados	Tasa de letalidad
1	2020-11-30	963624	36019	3.7032
2	2020-12-01	964891	36077	3.7027
3	2020-12-02	966118	36153	3.704
4	2020-12-03	967444	36232	3.7045
5	2020-12-04	968664	36307	3.704
6	2020-12-05	969883	36388	3.7038
7	2020-12-06	971330	36479	3.7028

ANEXO 3: OTROS ANEXOS

Desarrollo de la Metodología

Redes neuronales recurrentes para la predicción de casos por Covid-19 usando datos abiertos en el Perú, 2020

Marco de Trabajo SCRUM

Descripción del marco de trabajo

En este capítulo se establecerá el proceso de desarrollo del sistema a través de la metodología ágil SCRUM, planteando una visión general prosiguiendo de ello se detalla las fases que complementaron el desarrollo.

Se puede describir “Scrum ha dominado tanto el mercado que la gente tiene dificultades para hablar de Agile sin hacerlo específico de Scrum.”(Nader K. Rad, Frank Turley, 2018)

También “se describe como una estrategia de desarrollo flexible basada en un enfoque holístico, en el que todo el proceso es realizado por un equipo multifuncional en múltiples fases superpuestas, y el equipo El desarrollo trabaja para lograr un objetivo” (The Blokehead, 2017)

Visión general

Para el desarrollo del sistema se expondrán el lenguaje y herramientas que se utilizaron:

TABLA 10 SOFTWARE

Nº	Software
1	PyCharm Community Edition 2020.1.3
2	GitHub Desktop
3	Anaconda Navigator

TABLA 11 FRAMEWORK

N°	Framework
1	Django
2	Bootstrap

TABLA 12 LENGUAJE

N°	Lenguaje
1	Python
2	JavaScript

TABLA 13 LIBRERÍAS

N°	Librerías
1	scikit-learn==0.23.2
2	beautifulsoup4==4.9.1
3	Django==3.0.8
4	dj-database-url==0.5.0
5	h5py==2.10.0
6	gunicorn==20.0.4
7	numpy==1.19.0

8	Keras==2.3.1
9	scipy==1.4.1
10	pandas==1.1.0
11	plotly==4.11.0
12	psycpg2==2.8.6
13	python-decouple==3.3
14	requests==2.25.0
15	tensorflow==1.14.0
16	whitenoise==5.2.0

TABLA 14 HOSTING

N°	Hosting
1	Heroku

Visión del Sistema

- Creación de un algoritmo de una red neuronal para la predicción de casos confirmados y fallecidos de tal manera que los resultados obtenidos se pueda sacar la tasa de letalidad y mortalidad.

Identificar al Scrum Master y a los Stakeholders

Scrum Master. Se define que “El Scrum Master es responsable de promover y apoyar a Scrum como se define en la Guía de Scrum. Los Scrum Masters hacen esto ayudando a todos a comprender la teoría, las prácticas, las reglas y los valores de Scrum.” (Vishal Malhotra, 2020)

Development Team. Se define que “son expertos en el área de aplicaciones que son responsables de entregar los elementos de la lista de trabajos pendientes y administrar sus propios esfuerzos.[...] capaces de encontrar su propio camino en lugar de recibir pedidos. Deben estar alineados con el objetivo del proyecto en lugar de trabajar a ciegas. Una tarea puede ser asignado a un solo miembro durante el Sprint, pero todo el Equipo de Desarrollo seguirá siendo responsable de esa tarea, ningún individuo es dueño de ninguna tarea” (Nader K. Rad, Frank Turley, 2018)

TABLA 15 ROLES Y RESPONSABLES

Roles	Responsables
Scrum Master	Saulo Catunta Esquivel
Development Team	Dalyuska M. Roncal Hidalgo

Historia de Usuario

TABLA 16 HISTORIA DE USUARIO HU01

Historia de Usuario	
ID	HU01
Nombre	Realizar predicción de personas confirmadas con Covid 19
Prioridad	Alta
Riesgo	Alta
Descripción	El sistema realizará una predicción con los 7 últimos registros .de los casos
Validación	<ul style="list-style-type: none">Se visualizará la predicción de 7 días

TABLA 17 HISTORIA DE USUARIO HU02

Historia de Usuario	
ID	HU02
Nombre	Realizar predicción de personas fallecidas por Covid 19
Prioridad	Alta
Riesgo	Alta
Descripción	El sistema realizará una predicción con los 7 últimos registros de los casos
Validación	<ul style="list-style-type: none">Se visualizará la predicción de 7 días

TABLA 18 HISTORIA DE USUARIO HU03

Historia de Usuario	
ID	HU03
Nombre	Actualizar datos de los casos confirmados de Covid 19
Prioridad	Alta
Riesgo	Alta
Descripción	El sistema utilizará la data actualizada de la plataforma de datos abiertos
Validación	<ul style="list-style-type: none"> • Ir a la página de plataforma de datos abierto y corroborar que los datos de los casos confirmados estén actualizados

TABLA 19 HISTORIA DE USUARIO HU04

Historia de Usuario	
ID	HU04
Nombre	Actualizar datos de los casos fallecidos por Covid 19
Prioridad	Alta
Riesgo	Alta
Descripción	El sistema utilizará la data actualizada de la plataforma de datos abiertos
Validación	<ul style="list-style-type: none"> • Ir a la página de plataforma de datos abierto y corroborar que los datos de los casos fallecidos estén actualizados

TABLA 20 HISTORIA DE USUARIO HU05

Historia de Usuario	
ID	HU05
Nombre	Generar gráfico de serie de tiempo con los registros de casos confirmados por día con Covid 19
Prioridad	Alta
Riesgo	Alta
Descripción	El sistema realizará una gráfico de serie de tiempo que permita visualizar los casos de personas confirmados por día con Covid 19
Validación	<ul style="list-style-type: none"> ● Ingresar a la pestaña de predicción del sistema y ubicar el gráfico confirmados por día con Covid 19

TABLA 21 HISTORIA DE USUARIO HU06

Historia de Usuario	
ID	HU06
Nombre	Generar gráfico de serie de tiempo con los registros de casos confirmados acumulados con Covid 19
Prioridad	Alta
Riesgo	Alta
Descripción	El sistema realizará una gráfico de serie de tiempo que permita visualizar los casos de personas confirmados acumulados con Covid 19
Validación	<ul style="list-style-type: none"> ● Ingresar a la pestaña de predicción del sistema y ubicar el gráfico confirmados acumulados con Covid 19

TABLA 22 HISTORIA DE USUARIO HU07

Historia de Usuario	
ID	HU07
Nombre	Generar gráfico de serie de tiempo con los registros de casos fallecidos por día por el Covid 19
Prioridad	Alta
Riesgo	Alta
Descripción	El sistema realizará una gráfico de serie de tiempo que permita visualizar los casos de personas fallecidos por día por Covid 19
Validación	<ul style="list-style-type: none"> ● Ingresar a la pestaña de predicción del sistema y ubicar el gráfico de casos fallecidos por día por el Covid 19

TABLA 23 HISTORIA DE USUARIO HU08

Historia de Usuario	
ID	HU08
Nombre	Generar gráfico de serie de tiempo con los registros de casos fallecidos acumulado por el Covid 19
Prioridad	Alta
Riesgo	Alta
Descripción	El sistema realizará un gráfico de serie de tiempo que permita visualizar los casos de personas fallecidos acumulado por Covid 19
Validación	<ul style="list-style-type: none"> ● Ingresar a la pestaña de predicción del sistema y ubicar el gráfico de fallecidos acumulado por el Covid 19

TABLA 24 HISTORIA DE USUARIO HU09

Historia de Usuario	
ID	HU09
Nombre	Mostrar tabla de los casos confirmados por Covid-19
Prioridad	Alta
Riesgo	Alta
Descripción	Se visualizará los datos reales y de la predicción
Validación	<ul style="list-style-type: none"> ● Ingresar a la pestaña de predicción del sistema

TABLA 25 HISTORIA DE USUARIO HU10

Historia de Usuario	
ID	HU10
Nombre	Mostrar tabla de los casos fallecidos por Covid-19
Prioridad	Alta
Riesgo	Alta
Descripción	Se visualizará los datos reales y de la predicción
Validación	<ul style="list-style-type: none"> ● Ingresar a la pestaña de predicción del sistema

TABLA 26 HISTORIA DE USUARIO HU11

Historia de Usuario	
ID	HU11
Nombre	Mostrar tabla de Tasa de mortalidad de Covid-19 por cada 10 000 habitantes
Prioridad	Alta
Riesgo	Alta
Descripción	Se visualizará los datos reales y de la predicción.
Validación	<ul style="list-style-type: none"> • Ingresar a la pestaña de predicción del sistema

TABLA 27 HISTORIA DE USUARIO HU12

Historia de Usuario	
ID	HU12
Nombre	Mostrar tabla de la tasa de letalidad de Covid-19
Prioridad	Alta
Riesgo	Alta
Descripción	Se visualizará los datos reales y de la predicción
Validación	<ul style="list-style-type: none"> • Ingresar a la pestaña de predicción del sistema

TABLA 28 HISTORIA DE USUARIO HU13

Historia de Usuario	
ID	HU13
Nombre	Selección de departamento o Perú para predecir
Prioridad	Alta
Riesgo	Alta
Descripción	Al seleccionar en el botón se desplegará los departamentos del Perú y al país Perú en general para que se escoja de que lugar se quiere predecir.
Validación	<ul style="list-style-type: none"> • Seleccionar mediante el botón ubicado en el apartado Lugar.

TABLA 29 HISTORIA DE USUARIO HU14

Historia de Usuario	
ID	HU14
Nombre	Mostrar los casos totales de Covid--19 del MINSA
Prioridad	Media
Riesgo	Media
Descripción	Se visualizará el número total de casos confirmados, fallecidos, recuperados y pruebas.
Validación	<ul style="list-style-type: none"> • Ingresar a la pestaña de Inicio del sistema (Página principal)

TABLA 30 HISTORIA DE USUARIO HU15

Historia de Usuario	
ID	HU15
Nombre	Mostrar gráfico circular de los casos de Covid-19
Prioridad	Media
Riesgo	Media
Descripción	Se mostrará un gráfico que facilitará la comprensión de cómo van los casos hasta el momento.
Validación	<ul style="list-style-type: none"> ● Ingresar a la pestaña de Inicio del sistema (Página principal)

Pila del Producto (Product Backlog)

TABLA 31 PILA DE PRODUCTO

HISTORIA DE USUARIO		
ID	QUIÉN...	RAZÓN
HU01	Predicción	Realiza una predicción con los 7 últimos registros .de los casos
HU02	Predicción	Realiza una predicción con los 7 últimos registros de los casos
HU03	Administrador	Actualiza datos de los casos confirmados de Covid 19
HU04	Administrador	Actualiza datos de los casos fallecidos por Covid 19
HU05	Predicción	Genera gráfico de serie de tiempo con los registros de casos confirmados por día con Covid 19

HU06	Predicción	Genera gráfico de serie de tiempo con los registros de casos confirmados acumulados con Covid 19
HU07	Predicción	Genera gráfico de serie de tiempo con los registros de casos fallecidos por día por el Covid 19
HU08	Predicción	Realiza un gráfico de serie de tiempo que permita visualizar los casos de personas fallecidos acumulado por Covid 19
HU09	Predicción	Muestra tabla de los casos confirmados por Covid-19
HU10	Predicción	Muestra tabla de los casos fallecidos por Covid-19
HU11	Predicción	Muestra tabla de Tasa de mortalidad de Covid-19 por cada 10 000 habitantes
HU12	Predicción	Muestra tabla de la tasa de letalidad de Covid-19
HU13	Usuario	Selecciona de departamento o Perú para predecir
HU14	Usuario	Muestra los casos totales de Covid--19 del MINSA
HU15	Usuario	Muestra gráfico circular de los casos de Covid-19

Sprints

TABLA 32 SPRINT GENERAL

INICIO	30/09/20	NOMBRE		
FIN	30/11/20	CovidPredict		
SPRINT	HISTORIA DE USUARIO	Pendiente	En curso	Completo
SPRINT 1	HU01			
	HU02			
	HU03			
	HU04			
SPRINT 2	HU05			
	HU06			
	HU07			
	HU08			
SPRINT 3	HU09			
	HU10			
	HU11			
	HU12			
SPRINT 4	HU13			
	HU14			
	HU15			

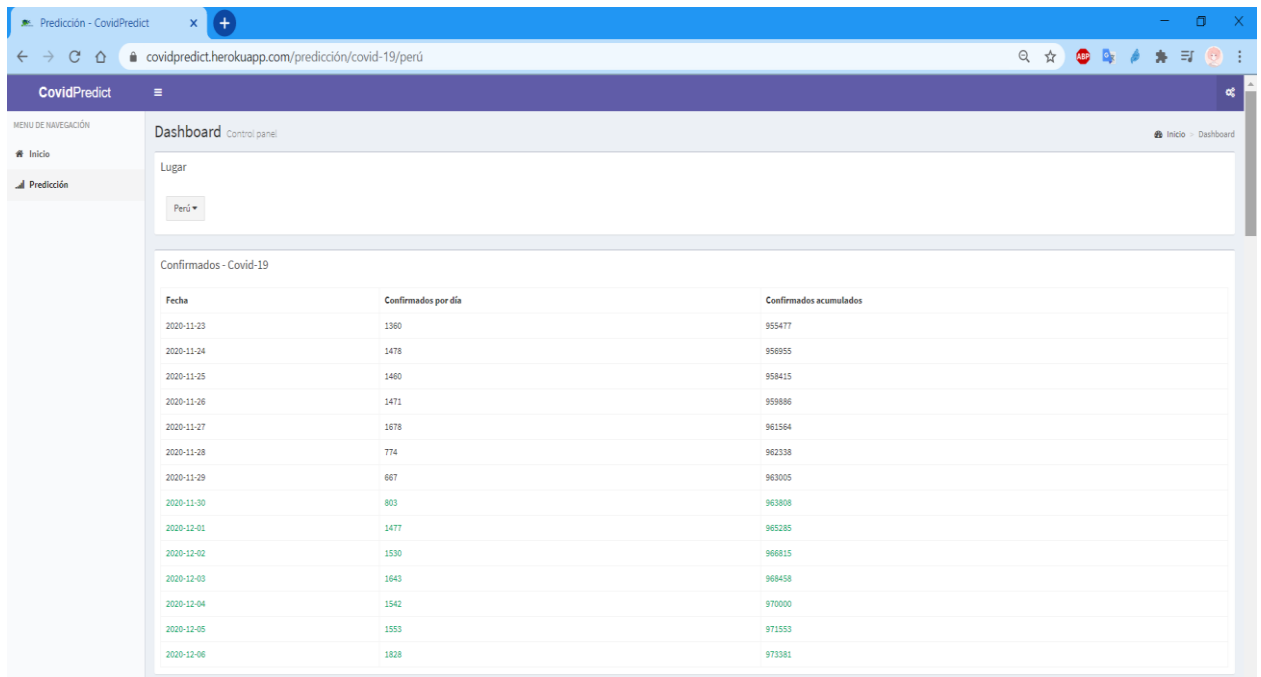
- ROJO --- INCOMPLETO
- NARANJA ----EN CURSO
- VERDE ---COMPLETO

TABLA 33 SPRINT 1 ENTRADA

INICIO	30/09/2020	NOMBRE		
FIN	14/10/2020	CovidPredict		
SPRINT	HISTORIA DE USUARIO	Pendiente	En curso	Completo
SPRINT 1	HU01 HU02 HU03 HU04			

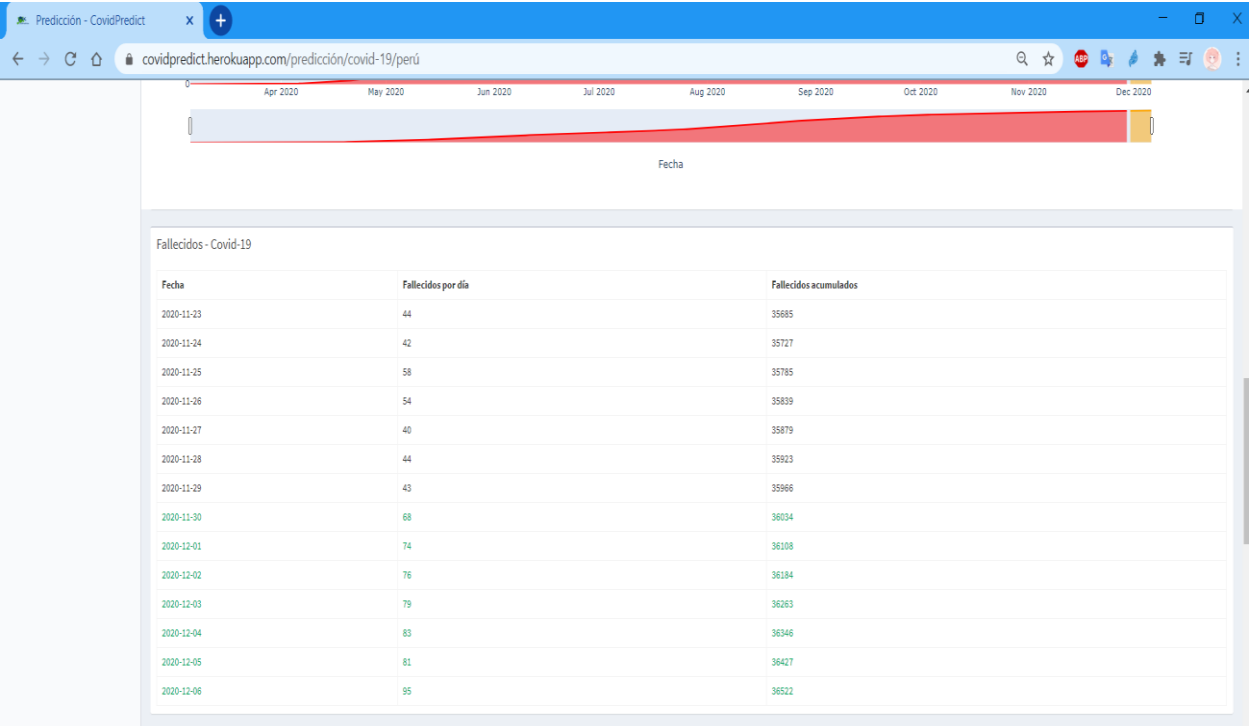
HU01

FIGURA 16 PREDICCIÓN DE PERSONAS CONFIRMADAS CON COVID 19



HU02

FIGURA 17 PREDICCIÓN DE PERSONAS FALLECIDAS POR COVID 19

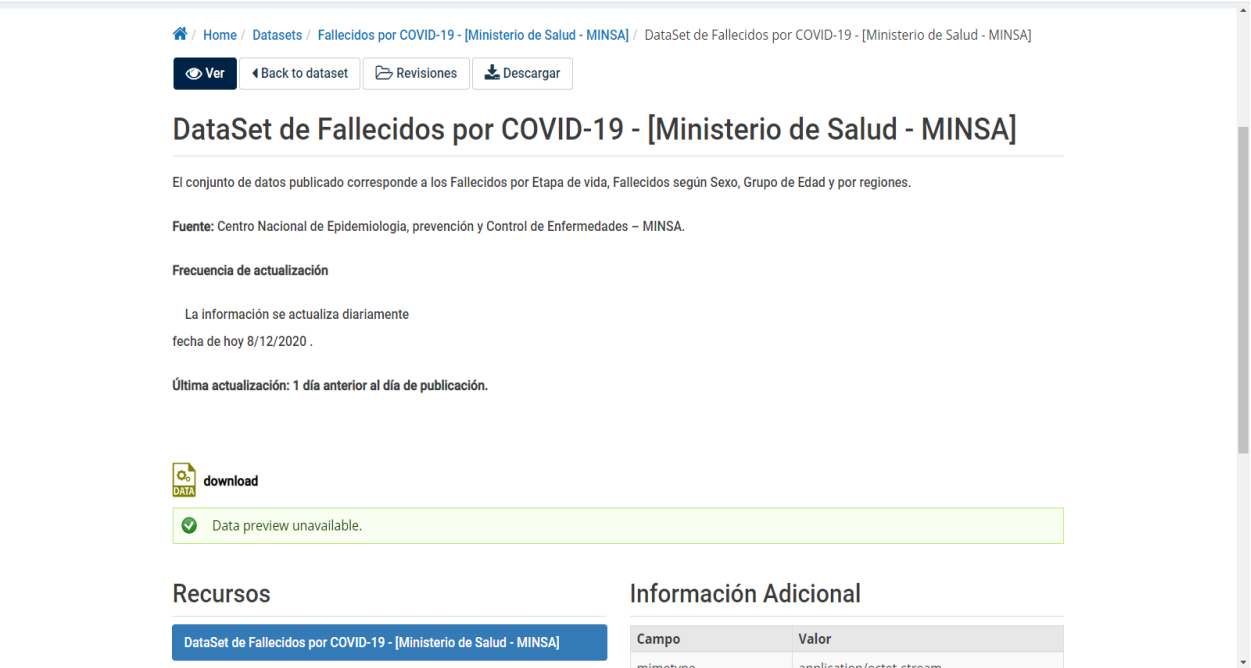


HU03

FIGURA 18 ACTUALIZAR DATOS DE LOS CASOS CONFIRMADOS DE COVID 19



FIGURA 19 ACTUALIZAR DATOS DE LOS CASOS FALLECIDOS POR COVID 19



Resumen del Sprint

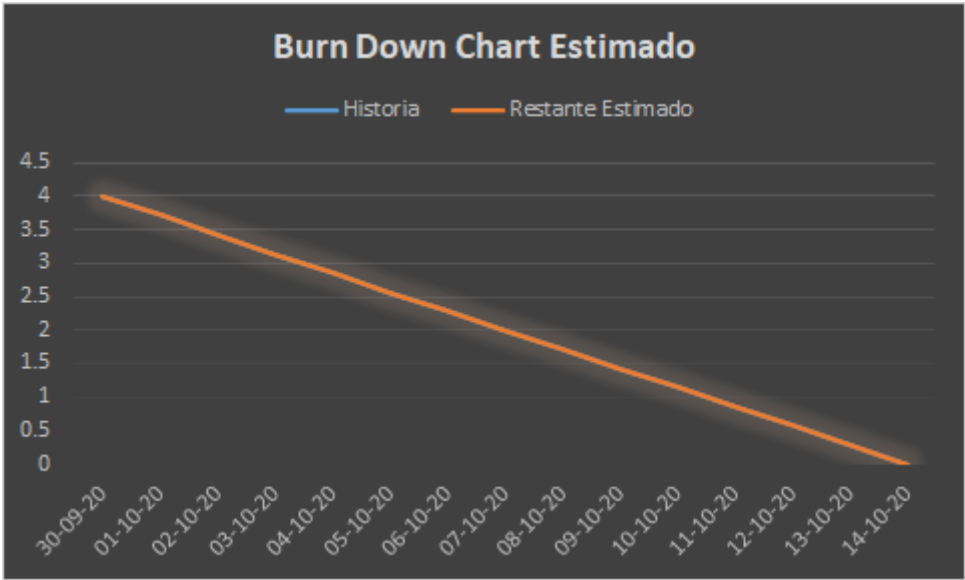
TABLA 34 RESUMEN SPRINT 1

ITEM	VALOR
Historias	4
Historias completas	4
Historias pendientes	0
Avance (%)	100%

Gráfico Burn down del Sprint 1

Estimado

FIGURA 20 BURN DOWN CHART ESTIMADO SPRINT 1



Real

FIGURA 21 BURN DOWN CHART REAL SPRINT 1

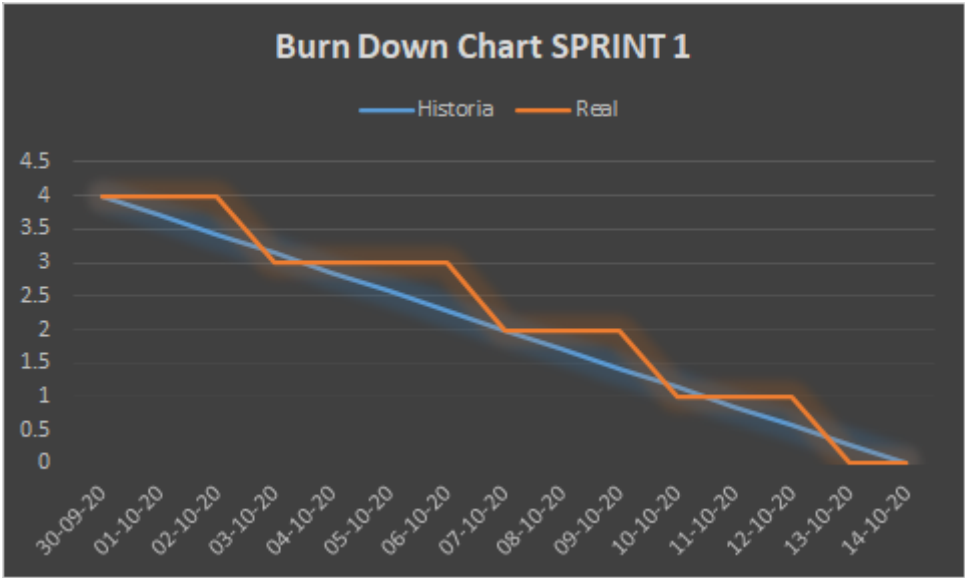


TABLA 35 SPRINT 1 SALIDA

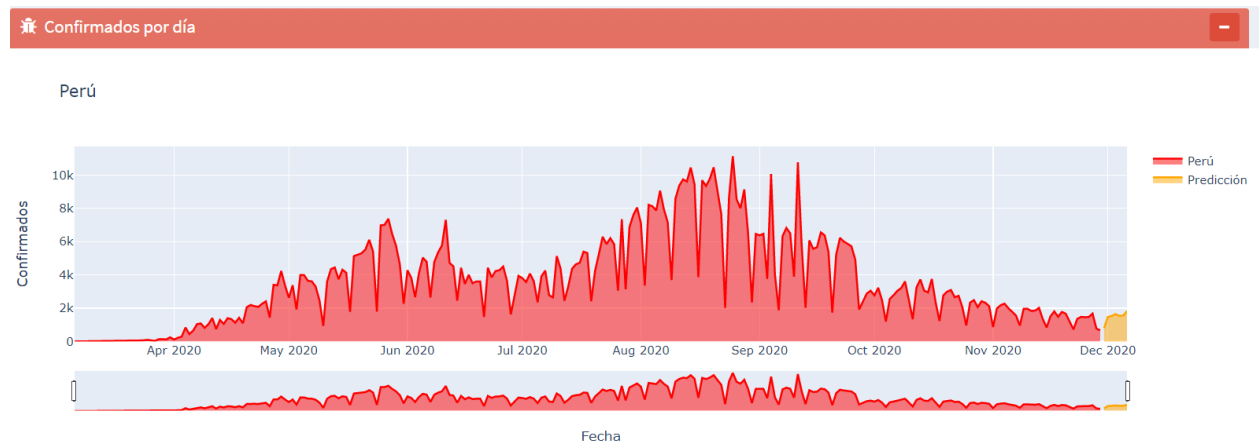
INICIO	30/09/2020	NOMBRE		
FIN	14/10/2020	CovidPredict		
SPRINT	HISTORIA DE USUARIO	Pendiente	En curso	Completo
SPRINT 1	HU01 HU02 HU03 HU04			

TABLA 36 SPRINT 2 ENTRADA

INICIO	15/10/2020	NOMBRE		
FIN	29/10/2020	CovidPredict		
SPRINT	HISTORIA DE USUARIO	Pendiente	En curso	Completo
SPRINT 1	HU01 HU02 HU03 HU04			
SPRINT 2	HU05 HU06 HU07 HU08			

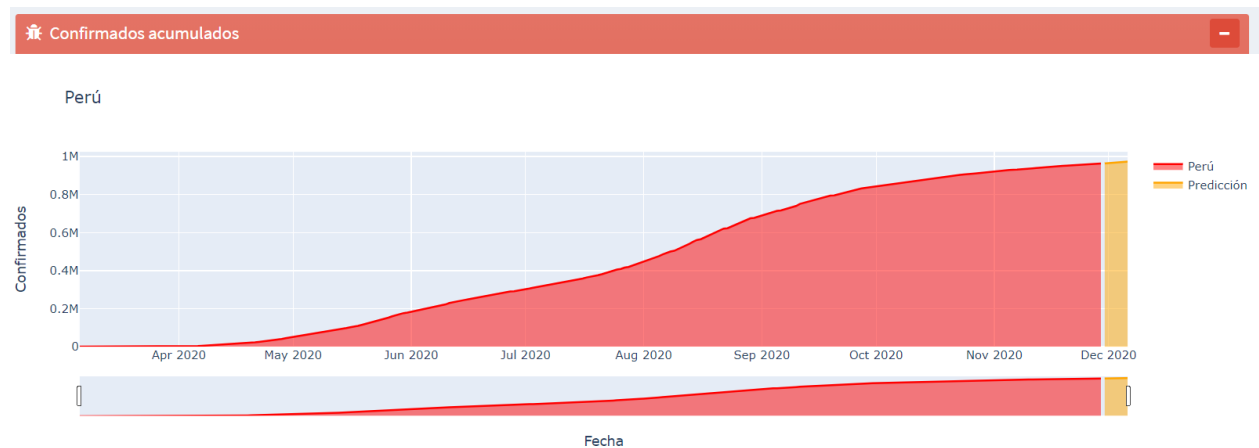
HU05

FIGURA 22 GENERAR GRÁFICO DE SERIE DE TIEMPO CON LOS REGISTROS DE CASOS CONFIRMADOS POR DÍA CON COVID 19



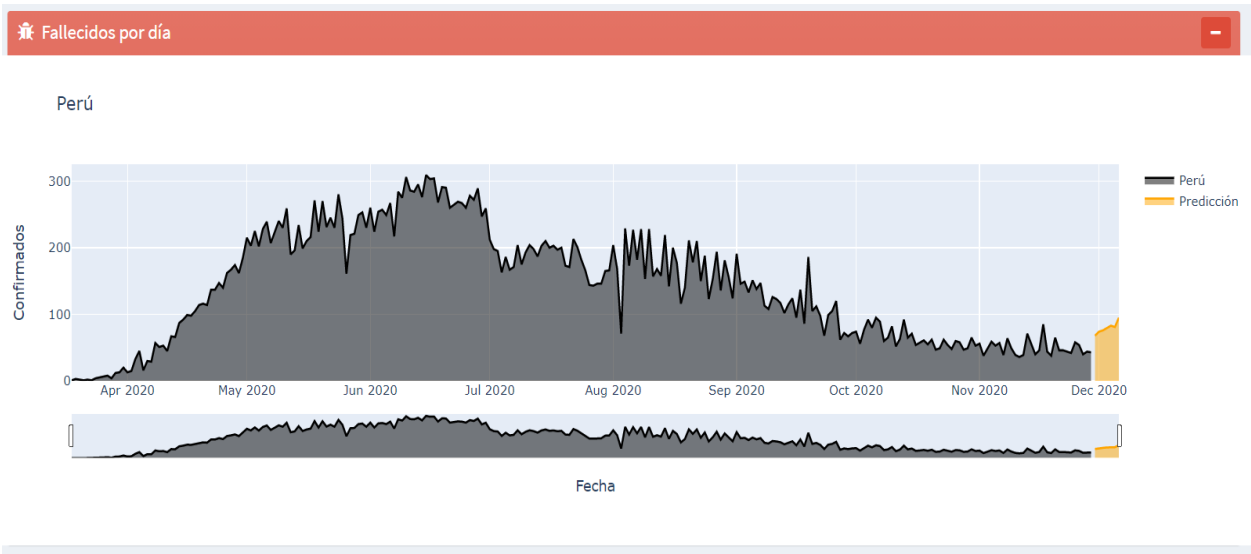
HU06

FIGURA 23 GENERAR GRÁFICO DE SERIE DE TIEMPO CON LOS REGISTROS DE CASOS CONFIRMADOS ACUMULADOS CON COVID 19



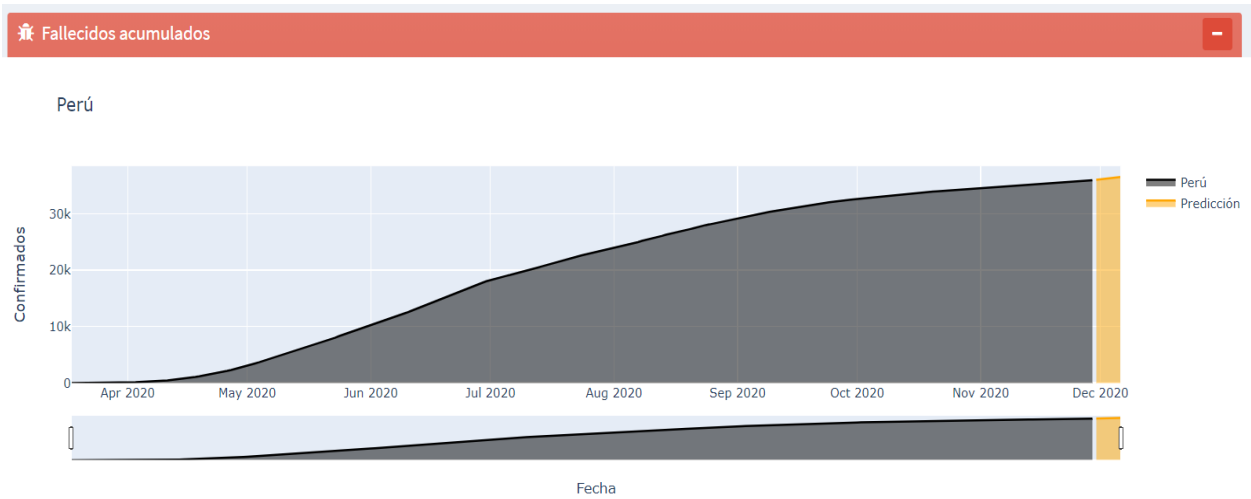
HU07

FIGURA 24 GENERAR GRÁFICO DE SERIE DE TIEMPO CON LOS REGISTROS DE CASOS FALLECIDOS POR DÍA POR COVID 19



HU08

FIGURA 25 GENERAR GRÁFICO DE SERIE DE TIEMPO CON LOS REGISTROS DE CASOS FALLECIDOS ACUMULADOS POR COVID 19



Resumen del Sprint

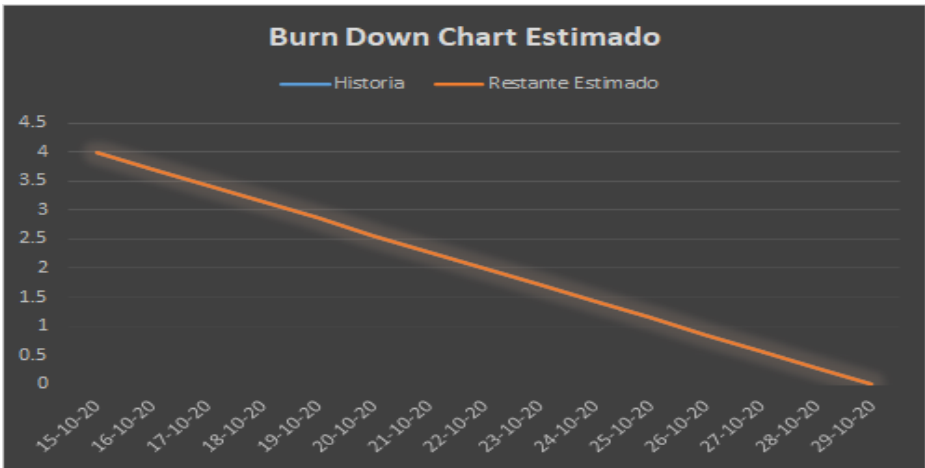
TABLA 37 RESUMEN SPRINT 2

ITEM	VALOR
Historias	4
Historias completas	4
Historias pendientes	0
Avance (%)	100%

Gráfico Burn down del Sprint 2

Estimado

FIGURA 26 BURN DOWN CHART ESTIMADO SPRINT 2



Real

FIGURA 27 BURN DOWN CHART REAL SPRINT 2

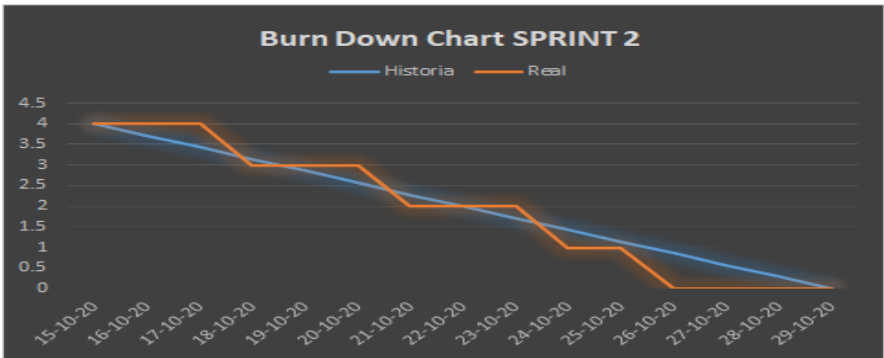


TABLA 38 SPRINT 2 SALIDA

INICIO	15/10/2020	NOMBRE		
FIN	29/10/2020	CovidPredict		
SPRINT	HISTORIA DE USUARIO	Pendiente	En curso	Completo
SPRINT 1	HU01			
	HU02			
	HU03			
	HU04			
SPRINT 2	HU05			
	HU06			
	HU07			
	HU08			

TABLA 39 SPRINT 3 ENTRADA

INICIO	30/10/2020	NOMBRE		
FIN	13/11/2020	CovidPredict		
SPRINT	HISTORIA DE USUARIO	Pendiente	En curso	Completo
SPRINT 1	HU01			
	HU02			
	HU03			
	HU04			
SPRINT 2	HU05			

	HU06 HU07 HU08			
SPRINT 3	HU09 HU10 HU11 HU12			

HU09

FIGURA 28 MOSTRAR TABLA DE LOS CASOS CONFIRMADOS CON COVID-19

Confirmados - Covid-19		
Fecha	Confirmados por día	Confirmados acumulados
2020-11-23	1360	955477
2020-11-24	1478	956955
2020-11-25	1460	958415
2020-11-26	1471	959886
2020-11-27	1678	961564
2020-11-28	774	962338
2020-11-29	667	963005
2020-11-30	803	963808
2020-12-01	1477	965285
2020-12-02	1530	966815
2020-12-03	1643	968458
2020-12-04	1542	970000
2020-12-05	1553	971553
2020-12-06	1828	973381

HU10

FIGURA 29 MOSTRAR TABLA DE LOS CASOS FALLECIDOS POR COVID-19

Fallecidos - Covid-19		
Fecha	Fallecidos por día	Fallecidos acumulados
2020-11-23	44	35685
2020-11-24	42	35727
2020-11-25	58	35785
2020-11-26	54	35839
2020-11-27	40	35879
2020-11-28	44	35923
2020-11-29	43	35966
2020-11-30	68	36034
2020-12-01	74	36108
2020-12-02	76	36184
2020-12-03	79	36263
2020-12-04	83	36346
2020-12-05	81	36427
2020-12-06	95	36522

HU11

FIGURA 30 MOSTRAR TABLA DE TASA DE MORTALIDAD DE COVID-19 POR CADA 10 000 HABITANTES

Tasa de mortalidad por cada 10 000 habitantes			
Fecha	Fallecidos acumulados	N° de habitantes	Tasa de mortalidad
2020-11-23	35685	32625948	10,94
2020-11-24	35727	32625948	10,95
2020-11-25	35785	32625948	10,97
2020-11-26	35839	32625948	10,98
2020-11-27	35879	32625948	11,0
2020-11-28	35923	32625948	11,01
2020-11-29	35966	32625948	11,02
2020-11-30	36034	32625948	11,04
2020-12-01	36108	32625948	11,07
2020-12-02	36184	32625948	11,09
2020-12-03	36263	32625948	11,11
2020-12-04	36346	32625948	11,14
2020-12-05	36427	32625948	11,17
2020-12-06	36522	32625948	11,19

HU12

FIGURA 31 MOSTRAR TABLA DE LA TASA DE LETALIDAD DE COVID-19

Tasa de letalidad			
Fecha	Confirmados acumulados	Fallecidos acumulados	Tasa de letalidad
2020-11-23	955477	35685	3.73 %
2020-11-24	956955	35727	3.73 %
2020-11-25	958415	35785	3.73 %
2020-11-26	959886	35839	3.73 %
2020-11-27	961564	35879	3.73 %
2020-11-28	962338	35923	3.73 %
2020-11-29	963005	35966	3.73 %
2020-11-30	963808	36034	3.7 %
2020-12-01	965285	36108	3.7 %
2020-12-02	966815	36184	3.7 %
2020-12-03	968458	36263	3.7 %
2020-12-04	970000	36346	3.7 %
2020-12-05	971553	36427	3.7 %
2020-12-06	973381	36522	3.69 %

Resumen del Sprint

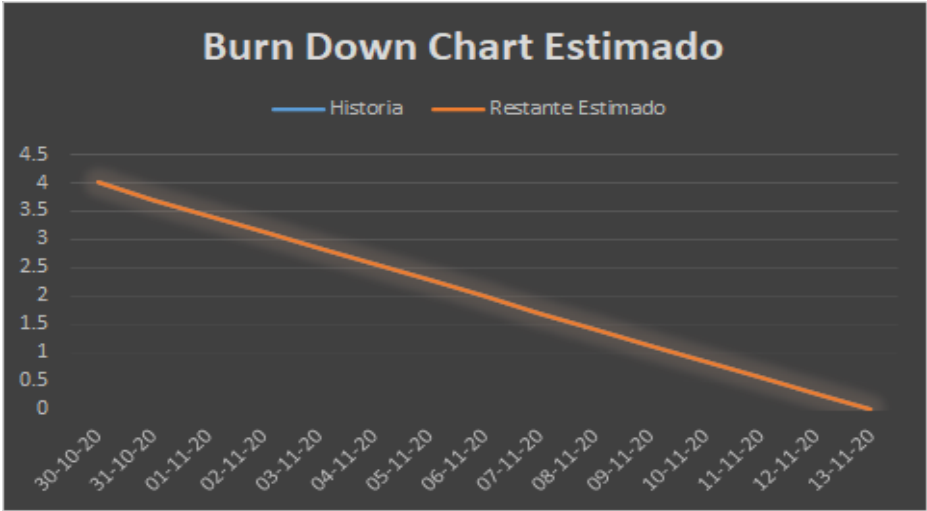
TABLA 40 RESUMEN SPRINT 3

ITEM	VALOR
Historias	4
Historias completas	4
Historias pendientes	0
Avance (%)	100%

Gráfico Burn down del Sprint 3

Estimado

FIGURA 32 BURN DOWN ESTIMADO SPRINT 3



Real

FIGURA 33 BURN DOWN CHART REAL SPRINT 3

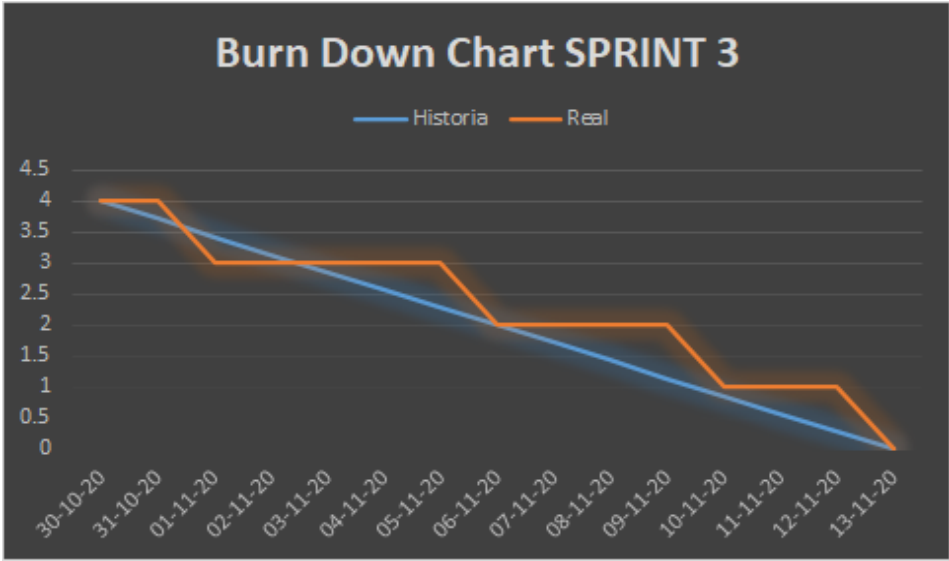


TABLA 41 SPRINT 3 SALIDA

INICIO	30/10/2020	NOMBRE		
FIN	13/11/2020	CovidPredict		
SPRINT	HISTORIA DE USUARIO	Pendiente	En curso	Completo
SPRINT 1	HU01			
	HU02			
	HU03			
	HU04			
SPRINT 2	HU05			
	HU06			
	HU07			
	HU08			
SPRINT 3	HU09			
	HU10			
	HU11			
	HU12			

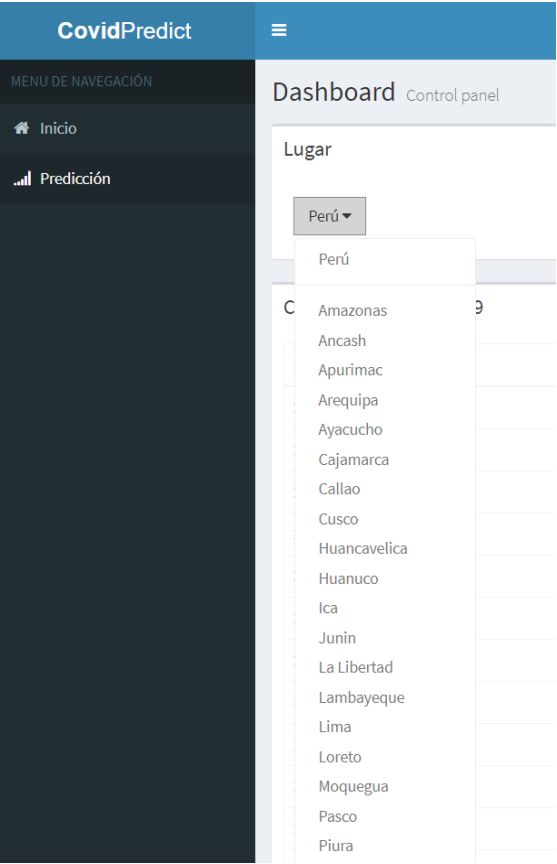
TABLA 42 SPRINT 4 ENTRADA

INICIO	16/11/2020	NOMBRE		
FIN	30/11/2020	CovidPredict		
SPRINT	HISTORIA DE USUARIO	Pendiente	En curso	Completo
SPRINT 1	HU01			
	HU02			

	HU03 HU04			
SPRINT 2	HU05 HU06 HU07 HU08			
SPRINT 3	HU09 HU10 HU11 HU12			
SPRINT 4	HU13 HU14 HU15			

HU13

FIGURA 34 SELECCIÓN DE DEPARTAMENTO O PERÚ PARA PREDECIR



HU14

FIGURA 35 MOSTRAR LOS CASOS TOTALES DE COVID--19 DEL MINSA



FIGURA 36 MOSTRAR GRÁFICO CIRCULAR DE LOS CASOS DE COVID-19



Resumen del Sprint

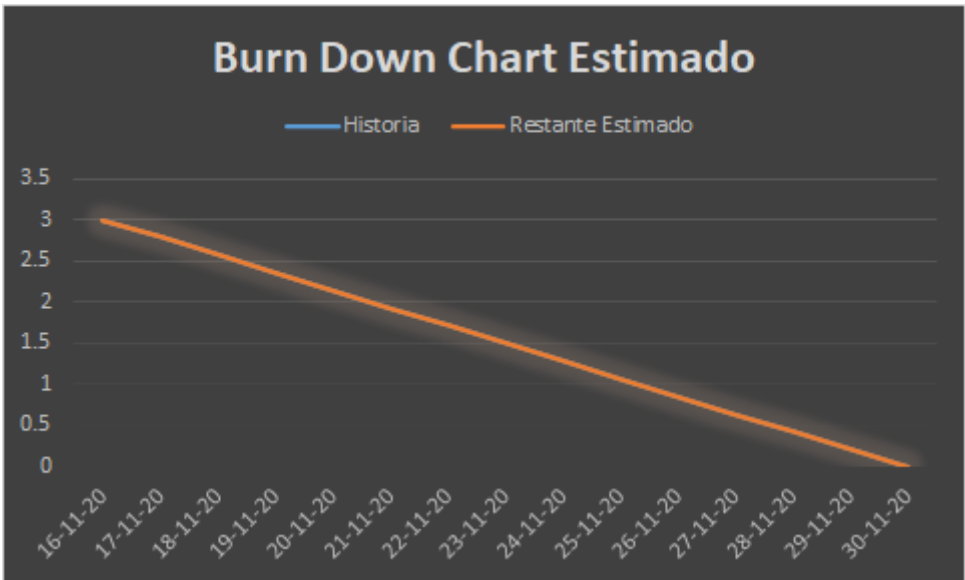
TABLA 43 RESUMEN SPRINT 4

ITEM	VALOR
Historias	3
Historias completas	4
Historias pendientes	0
Avance (%)	100%

Gráfico Burn down del Sprint 4

Estimado

FIGURA 37 BURN DOWN CHART ESTIMADO SPRINT 4



Real

FIGURA 38 BURN DOWN CHART REAL SPRINT 4

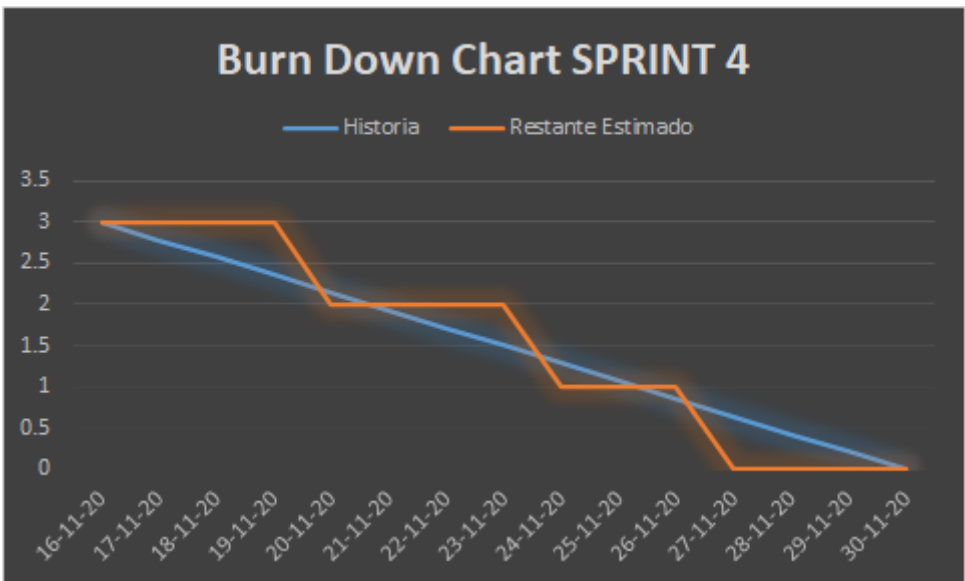
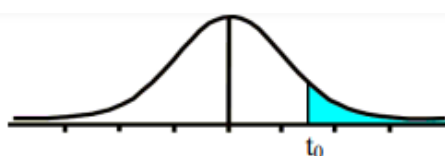


TABLA 44 SPRINT 4 SALIDA

INICIO	30/09/20	NOMBRE		
FIN	30/11/20	CovidPredict		
SPRINT	HISTORIA DE USUARIO	Pendiente	En curso	Completo
SPRINT 1	HU01 HU02 HU03 HU04			
SPRINT 2	HU05 HU06 HU07 HU08			
SPRINT 3	HU09 HU10 HU11 HU12			
SPRINT 4	HU13 HU14 HU15			

Tabla T-Student

Tabla t-Student



Grados de libertad	0.25	0.1	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.8165	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.7407	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7970
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
--	----	----	----	----	----	----

Selección de metodología por juicio de expertos

TABLA 45 SELECCIÓN DE METODOLOGÍA POR JUICIO DE EXPERTOS

*	METODOLOGÍA		
	SCRUM	XP	KANBAN
EXPERTOS			
EVEN	35	31	31
RODRIGUEZ	32	20	25
LIENDO	35	0	0
SUMATO- RIA	102	51	56
TOTAL	209		
PORCEN- TAJE	49%	24%	27%

CRISP-DM

1. Comprensión del Negocio

1.1. Determinar los objetivos del negocio

El objetivo de la minería de datos que se plantea en la investigación es el obtener predicciones confiables utilizando los datos de la plataforma de Datos Abiertos del Perú. El objetivo es predecir los casos por Covid-19.

Objetivos del negocio

- Predecir los casos confirmados por Covid-19 en el Perú.
- Predecir los casos fallecidos por Covid-19 en el Perú.
- Calcular el aumento de la tasa de mortalidad por Covid-19 en el Perú.
- Calcular el aumento la variación de la tasa de letalidad por Covid-19 en el Perú.

1.2. Evaluar la situación

Para trabajar la data para los casos por Covid-19 se tiene archivos en formato csv con información sobre casos contagio y fallecidos en el Perú y sus departamentos desde el 06-03-2020 hasta 29-11-2020 para los casos de contagios y para fallecidos es de 18-03-2020 hasta 29-11-2020. La información que será útil al momento de hacer la minería de datos que se encuentran en el formato csv serian departamento y la fecha de resultados.

1.3. Determinar el objetivo de Minería de Datos

Se determina el objetivo de la minería de datos son:

Predecir la cantidad de personas contagiadas de Covid-19 por día en el Perú.

Predecir la cantidad de personas contagiadas de Covid-19 acumulados en el Perú.

Predecir la cantidad de personas fallecidos por Covid-19 por día en el Perú.

Predecir la cantidad de personas fallecidos por Covid-19 acumulados en el Perú.

1.4. Plan de proyecto

Se divide en las siguientes etapas:

ETAPA	DESCRIPCIÓN	TIEMPO ESTIMADO
Etap 1	Comprensión de los objetivos del negocio	1 semana
Etap 2	Análisis de los datos y la información que conforma la base de datos.	1 semana
Etap 3	Preparación de datos para la elaboración de la minería de datos.	3 semanas
Etap 4	Modelamiento de datos.	2 semana

2. Comprensión de los Datos

2.1. Obtener los datos iniciales

Se utilizó la librería panda de Python para extraer los datos de la plataforma de datos abiertos del Perú.

En la Figura 39 se muestran los datos del archivo csv de casos confirmados por Covid-19, el cual contiene 9 campos con los siguientes nombres: FECHA_CORTE, UUID, DEPARTAMENTO, PROVINCIA, DISTRITO, METODODX, EDAD, SEXO y FECHA_RESULTADO, con una cantidad de 963605 registros.

FIGURA 39 CASOS CONFIRMADOS

	FECHA_CORTE	UUID	DEPARTAMENTO	PROVINCIA	DISTRITO	METODODX	EDAD	SEXO	FECHA_RESULTADO
0	20201129	7320cabdc1aaca5c59014cae76a134e6	LIMA	LIMA	VILLA MARIA DEL TRIUNFO	PCR	36.0	MASCULINO	20200802.0
1	20201129	e81602051997ace8340bb8c18fe24c65	LIMA	EN INVESTIGACIÓN	EN INVESTIGACIÓN	PCR	36.0	MASCULINO	20200803.0
2	20201129	cecdbf10074dbc011ae05b3cbd320a6f	LIMA	EN INVESTIGACIÓN	EN INVESTIGACIÓN	PCR	36.0	MASCULINO	20200803.0
3	20201129	71ecb6cccb248b0bb2ac72ed51b5e979	LIMA	EN INVESTIGACIÓN	EN INVESTIGACIÓN	PCR	36.0	MASCULINO	20200803.0
4	20201129	566af4276cbe9359abe93f9aa86396c3	LIMA	LIMA	SAN JUAN DE LURIGANCHO	PCR	36.0	MASCULINO	20200803.0
...
963600	20201129	cbba30be663ec2f041f5a1c8a81323b5	AMAZONAS	CONDORCANQUI	NIEVA	PR	42.0	MASCULINO	20200822.0
963601	20201129	9b26bce5dc7e5bd86f06d4148f3b23c8	LORETO	MARISCAL RAMON CASTILLA	RAMON CASTILLA	PR	43.0	MASCULINO	20200826.0
963602	20201129	48ddb6d6691a40199fb3575dc951bea3	LORETO	ALTO AMAZONAS	YURIMAGUAS	PR	43.0	MASCULINO	20200827.0
963603	20201129	5c3bdf415fadd15a82408478dee5b6f	LORETO	DATEM DEL MARAÑON	BARRANCA	PR	42.0	MASCULINO	20200813.0
963604	20201129	dd9a8aaa19c939f77368cedf34b6bc9	SAN MARTIN	PICOTA	SAN HILARION	PR	43.0	MASCULINO	20200818.0

963605 rows x 9 columns

En la Figura 40 se muestran los datos del archivo csv de casos fallecidos por Covid-19, el cual contiene 9 campos con los siguientes nombres:

FECHA_CORTE, UUID, FECHA_FALLECIMIENTO, EDAD_DECLARADA, SEXO, FECHA_NAC, DEPARTAMENTO, PROVINCIA, DISTRITO, con una cantidad de 35966 registros.

FIGURA 40 CASOS FALLECIDOS

	FECHA_CORTE	UUID	FECHA_FALLECIMIENTO	EDAD_DECLARADA	SEXO	FECHA_NAC	DEPARTAMENTO	PROVINCIA	DISTRITO
0	20201128	7320cabdc1aaca6c59014cae76a134e6	20200507	61	MASCULINO	NaN	ANCASH	NaN	CHIMBOTE
1	20201128	e81602051997ace8340bb8c18fe24c65	20200411	66	MASCULINO	19530708.0	LIMA	LIMA	LURIN
2	20201128	cecdbf10074dbc011ae05b3cbd320a6f	20200511	62	FEMENINO	19570902.0	LIMA	LIMA	LIMA
3	20201128	71ecb6bccb248b0bb2ac72ed51b5e979	20200422	59	FEMENINO	19601019.0	LAMBAYEQUE	FERREÑAFE	FERREÑAFE
4	20201128	566af4276cbe9359abe93f9aa86396c3	20200523	72	MASCULINO	NaN	ANCASH	NaN	CHIMBOTE
...
35961	20201128	1affbde055851efc6498a201d150d71e	20201101	52	MASCULINO	NaN	ICA	NAZCA	NAZCA
35962	20201128	f985df532d806f1ea2845aa19f1091bb	20201101	54	MASCULINO	NaN	ICA	NAZCA	VISTA ALEGRE
35963	20201128	6ae266266b3410ccc9821ab03d3a6a4e	20201101	59	FEMENINO	NaN	JUNIN	HUANCAYO	CHILCA
35964	20201128	be42ac545354477e46a6ea7aa74197ef	20201101	63	MASCULINO	NaN	JUNIN	CHANCHAMAYO	PERENE
35965	20201128	3597b50c825faf31a9803870114c2178	20201101	66	MASCULINO	NaN	LA LIBERTAD	TRUJILLO	LA ESPERANZA

35966 rows x 9 columns

2.2. Describir los datos

Descripción de los datos de confirmados por Covid-19

FIGURA 41 DESCRIPCIÓN DATOS CONFIRMADOS

	FECHA_CORTE	EDAD	FECHA_RESULTADO
count	963605	963569	963005
mean	20201129	42	20200761
std	0	18	179
min	20201129	0	20200306
25%	20201129	29	20200614
50%	20201129	41	20200806
75%	20201129	54	20200907
max	20201129	120	20201129

FIGURA 42 DETALLE DATOS CONFIRMADOS

FECHA_CORTE	963605
UUID	963605
DEPARTAMENTO	963605
PROVINCIA	963605
DISTRITO	963605
METODODX	963605
EDAD	963569
SEXO	963605
FECHA_RESULTADO	963005

De acuerdo a los resultados de la descripción de los datos se observa que los campos EDAD y FECHA_RESULTADO tiene datos vacíos.

Descripción de los datos de fallecidos por Covid-19

FIGURA 43 DESCRIPCIÓN DE LOS DATOS DE FALLECIDOS

	FECHA_CORTE	FECHA_FALLECIMIENTO	EDAD_DECLARADA	FECHA_NAC
count	35966	35966	35966	31748
mean	20201128	20200696	65	19536481
std	0	182	14	143313
min	20201128	20200318	0	19120319
25%	20201128	20200528	57	19431228
50%	20201128	20200630	67	19530421
75%	20201128	20200819	76	19621006
max	20201128	20201129	108	20200820

FIGURA 44 DETALLE DATOS DE FALLECIDOS

FECHA_CORTE	35966
UUID	35966
FECHA_FALLECIMIENTO	35966
EDAD_DECLARADA	35966
SEXO	35966
FECHA_NAC	31748
DEPARTAMENTO	35966
PROVINCIA	35115
DISTRITO	35645

De acuerdo a los resultados de la descripción de los datos fallecidos por Covid_19, se observa que los campos FECHA_NAC, PROVINCIA y DISTRITO tienen datos vacíos.

3. Preparación de los Datos

3.1. Seleccionar los datos

De los datos de confirmados por Covid-19 se ha seleccionado las columnas de DEPARTAMENTO Y FECHA_RESULTADO.

FIGURA 45 CONFIRMADOS POR COVID-19

	DEPARTAMENTO	FECHA_RESULTADO
0	LIMA	20200802.0
1	LIMA	20200803.0
2	LIMA	20200803.0
3	LIMA	20200803.0
4	LIMA	20200803.0
...
963600	AMAZONAS	20200822.0
963601	LORETO	20200826.0
963602	LORETO	20200827.0
963603	LORETO	20200813.0
963604	SAN MARTIN	20200818.0

963005 rows × 2 columns

De los datos de confirmados por Covid-19 se ha seleccionado las columnas de DEPARTAMENTO Y FECHA_FALLECIMIENTO.

FIGURA 46 DATOS DE CONFIRMADOS POR COVID-19 SELECCIÓN DE COLUMNAS DE DATOS CONFIRMADOS

	DEPARTAMENTO	FECHA_FALLECIMIENTO
0	ANCASH	20200507
1	LIMA	20200411
2	LIMA	20200511
3	LAMBAYEQUE	20200422
4	ANCASH	20200523
...
35961	ICA	20201101
35962	ICA	20201101
35963	JUNIN	20201101
35964	JUNIN	20201101
35965	LA LIBERTAD	20201101

35966 rows x 2 columns

3.2. Limpiar datos

En los datos de casos confirmados se encontraron 600 registros, en estos casos se pueden eliminar los registros o también ingresar un dato, quedando así a elección del investigador. No se ha visto viable inventar fechas, así que se tuvieron que eliminar los 600 registros vacíos correspondiente a la columna de FECHA_RESULTADO. Y para los casos de fallecidos no se encontraron registros vacíos, por lo que no se eliminó ningún registro.

3.3. Dar formato a los datos

En los datos confirmados, el campo FECHA_RESULTADO no tenía un formato de fecha, por lo que se realizó un formateo para convertirlo en fecha. El mismo proceso se realizó para el campo FECHA_FALLECIMIENTO de los datos de fallecidos.

3.4. Procesamiento de datos

Para los datos de confirmados por Covid-19 se realizó un conteo de la cantidad de registros por DEPARTAMENTO y FECHA_RESULTADO para obtener la cantidad de casos para cada fecha y departamento.

FIGURA 47 CANTIDAD DE CASOS PARA CADA FECHA Y DEPARTAMENTO

DEPARTAMENTO	AMAZONAS	ANCASH	APURIMAC	AREQUIPA	AYACUCHO	CAJAMARCA	CALLAO	CUSCO	HUANCAVELICA	HUANUCO	ICA	JUNIN	LA LIBERTAD	LAMBAYEQUE	LIMA	RE
FECHA																
2020-03-06	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
2020-03-07	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0
2020-03-08	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0
2020-03-09	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	1.0
2020-03-10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
...
2020-11-25	20.0	65.0	19.0	34.0	39.0	40.0	86.0	58.0	7.0	32.0	17.0	95.0	64.0	43.0	492.0	
2020-11-26	8.0	86.0	23.0	70.0	12.0	32.0	118.0	48.0	6.0	51.0	16.0	80.0	74.0	75.0	427.0	
2020-11-27	16.0	93.0	43.0	84.0	36.0	34.0	83.0	86.0	21.0	41.0	10.0	50.0	75.0	117.0	526.0	
2020-11-28	8.0	22.0	9.0	32.0	24.0	15.0	42.0	22.0	5.0	36.0	10.0	58.0	25.0	12.0	237.0	
2020-11-29	26.0	56.0	1.0	22.0	4.0	17.0	29.0	9.0	2.0	23.0	10.0	10.0	16.0	12.0	334.0	

268 rows x 27 columns

Luego se sumaron los casos de cada departamento para obtener los casos de Covid-19 del Perú.

FIGURA 48 CASOS DE COVID-19 DEL PERÚ

PERÚ	
FECHA	
2020-03-06	1.0
2020-03-07	5.0
2020-03-08	2.0
2020-03-09	3.0
2020-03-10	1.0
...	...
2020-11-25	1460.0
2020-11-26	1471.0
2020-11-27	1678.0
2020-11-28	774.0
2020-11-29	667.0

268 rows x 1 columns

Para los datos de fallecidos por Covid-19 se realizó un conteo de la cantidad de registros por DEPARTAMENTO y FECHA_FALLECIMIENTO para obtener la cantidad de casos para cada fecha y departamento.

FIGURA 49 CANTIDAD DE CASOS FALLECIDOS POR COVID-19

DEPARTAMENTO	AMAZONAS	ANCASH	APURIMAC	AREQUIPA	AYACUCHO	CAJAMARCA	CALLAO	CUSCO	HUANCANELICA	HUANUCO	ICA	JUNIN	LA LIBERTAD	LAMBAYEQUE	LIMA	LOR
FECHA																
2020-03-18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
2020-03-19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0
2020-03-20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0
2020-03-21	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2020-03-22	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...
2020-11-25	0.0	3.0	1.0	0.0	1.0	2.0	2.0	1.0	0.0	0.0	5.0	1.0	3.0	2.0	24.0	
2020-11-26	1.0	3.0	0.0	4.0	0.0	1.0	4.0	1.0	0.0	1.0	2.0	2.0	2.0	3.0	18.0	
2020-11-27	0.0	2.0	0.0	0.0	1.0	0.0	4.0	1.0	0.0	3.0	0.0	3.0	5.0	4.0	12.0	
2020-11-28	0.0	2.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	2.0	2.0	0.0	6.0	22.0	
2020-11-29	0.0	1.0	1.0	1.0	0.0	0.0	4.0	2.0	0.0	1.0	0.0	2.0	0.0	1.0	22.0	

256 rows x 26 columns

Luego se sumaron los casos de cada departamento para obtener los casos de Covid-19 del Perú.

FIGURA 50 CASOS DE COVID-19 DEL PERÚ- FALLECIDOS

PERÚ	
FECHA	
2020-03-18	1.0
2020-03-19	3.0
2020-03-20	2.0
2020-03-21	1.0
2020-03-22	2.0
...	...
2020-11-25	58.0
2020-11-26	54.0
2020-11-27	40.0
2020-11-28	44.0
2020-11-29	43.0

256 rows x 1 columns

4. Modelado

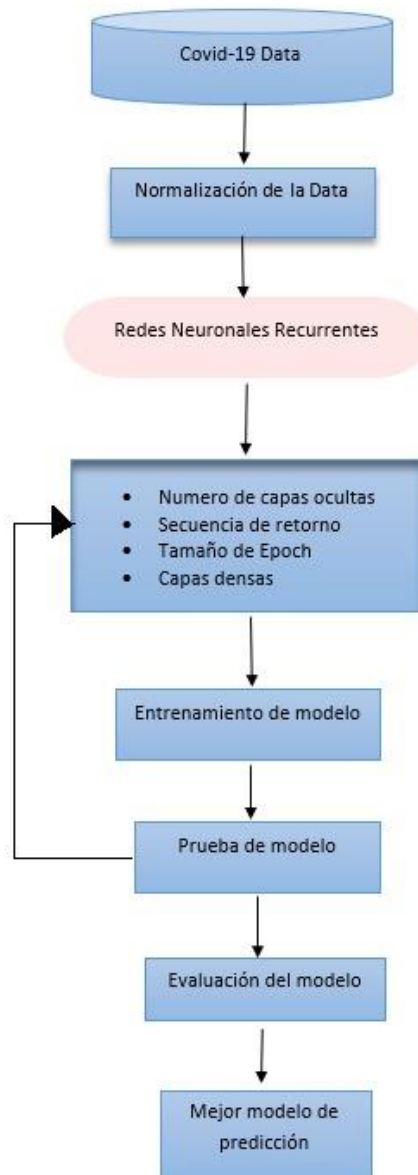
4.1. Seleccionar técnica de modelamiento

Para normalizar los datos se utilizó el método “MinMaxScaler()” de la librería Sklearn. De esta manera los datos se normalizaron a una escala de 0 a 1.

4.2. Generar el diseño de prueba

Se utilizaron varios tipos de redes neuronales recurrentes como LSTM, GRU, BI-LSTM y BI-GRU.

FIGURA 51 DISEÑO DE PRUEBA



4.3. Construir el modelo

Para entrenar la red neuronal recurrente se realizó una separación de los datos en un 80% para entrenamiento y 20% para prueba. Y para la validación se utilizó el parámetro “validation_split” del método “fit()” de tensorflow y keras, el cual recibe el valor de 0.2 que equivale al 20% de los datos de entrenamiento.

4.4. Evaluar el modelo

Se realizó el entrenamiento del modelo con 100 épocas para cada RNN de confirmados y fallecidos.

TABLA 46 FALLECIDOS COVID-19 - PREDICCIÓN

Fallecidos por Covid-19 - Predicción de 7 días (RMSE)								
Arquitectura	RNN	RMSE 1	RMSE 2	RMSE 3	RMSE 4	RMSE 5	Media	Valor mínimo
Simple	LSTM	0.1454	0.12907	0.13786	0.1273	0.16307	0.14054	0.12730
	GRU	0.15594	0.1425	0.14118	0.14895	0.10952	0.13962	0.10952
	BI-LSTM	0.12797	0.13733	0.13375	0.11114	0.12785	0.12761	0.11114
	BI-GRU	0.13503	0.10917	0.09672	0.11034	0.09221	0.10869	0.09221
Stacked	LSTM - LSTM	0.16898	0.15699	0.11906	0.15883	0.15581	0.15193	0.11906
	GRU - GRU	0.15611	0.12983	0.12567	0.11813	0.13008	0.13196	0.11813
	BI-LSTM - BI-LSTM	0.11535	0.12246	0.1313	0.12789	0.12082	0.12356	0.11535
	BI-GRU - BI-GRU	0.11278	0.11158	0.1166	0.10816	0.12529	0.11488	0.10816
Hibrido	LSTM - GRU	0.1379	0.16399	0.18569	0.17367	0.13717	0.15968	0.13717
	GRU - LSTM	0.16263	0.15478	0.13131	0.14452	0.12813	0.14427	0.12813
	BI-LSTM - BI-GRU	0.14071	0.13841	0.12477	0.13954	0.12969	0.13462	0.12477
	BI-GRU - BI-LSTM	0.09401	0.09384	0.11531	0.08606	0.11481	0.10081	0.08606

FIGURA 52 ENTRENAMIENTO DE LA RED NEURONAL RECURRENTE PARA FALLECIDOS



FIGURA 53 RESULTADO RMSE PARA FALLECIDOS

```
[405] rmse_7dias = math.sqrt(mean_squared_error(y_test[len(y_test)-semanas*serie_tiempo:], _test_predict[len(_test_predict)-serie_tiempo*semanas:]))
print('Error cuadrático medio (RMSE): %f' % (rmse_7dias))

Error cuadrático medio (RMSE): 0.086056
```

FIGURA 54 CONFIRMADOS CON COVID-19 - PREDICCIÓN

Confirmados por Covid-19 - Predicción de 7 días (RMSE)								
Arquitectura	RNN	RMSE 1	RMSE 2	RMSE 3	RMSE 4	RMSE 5	Media	Valor mínimo
Simple	LSTM	0.06005	0.05906	0.07131	0.05701	0.05607	0.06070	0.05607
	GRU	0.0568	0.058	0.07275	0.06341	0.05928	0.06205	0.05680
	BI-LSTM	0.05695	0.05607	0.06457	0.06419	0.06298	0.06095	0.05607
	BI-GRU	0.07069	0.05997	0.07345	0.0579	0.05416	0.06323	0.05416
Stacked	LSTM - LSTM	0.06377	0.05718	0.0563	0.05634	0.06494	0.05971	0.05630
	GRU - GRU	0.07296	0.07307	0.07548	0.05731	0.05731	0.06723	0.05731
	BI-LSTM - BI-LSTM	0.05294	0.05182	0.06049	0.0531	0.05547	0.05476	0.05182
	BI-GRU - BI-GRU	0.0692	0.04844	0.05751	0.0526	0.06282	0.05811	0.04844
Hibrido	LSTM - GRU	0.06457	0.06746	0.05483	0.0547	0.06114	0.06054	0.05470
	GRU - LSTM	0.06202	0.05791	0.06512	0.0576	0.05875	0.06028	0.05760
	BI-LSTM - BI-GRU	0.06427	0.05651	0.05949	0.06564	0.05576	0.06033	0.05576
	BI-GRU - BI-LSTM	0.044502	0.05955	0.05492	0.05836	0.0688	0.05723	0.04450

FIGURA 55 ENTRENAMIENTO DE LA RED NEURONAL RECURRENTE PARA CONFIRMADOS

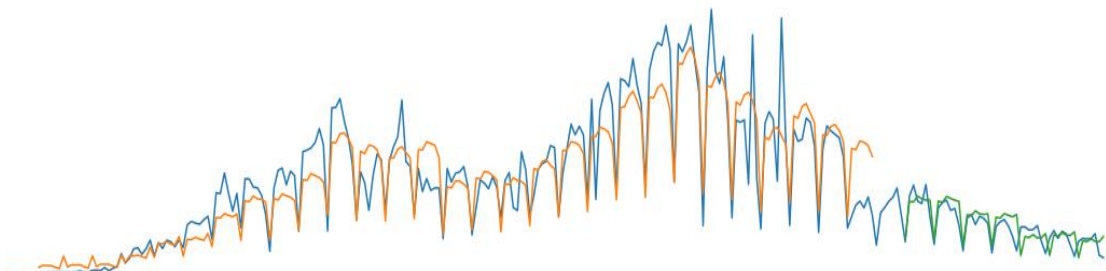


FIGURA 56 RESULTADOS RMSE CONFIRMADOS

```
[2110] rmse_7dias = math.sqrt(mean_squared_error(_y_test[len(_y_test)-semanas*serie_tiempo:], _test_predict[len(_test_predict)-serie_tiempo*semanas:]))
print('Error cuadrático medio (RMSE): %f' % (rmse_7dias))

Error cuadrático medio (RMSE): 0.044502
```

De acuerdo a los resultados obtenidos se utilizó la arquitectura híbrida BI-GRU – BI-LSTM el cual dio un valor mínimo de RMSE con 0,04450 para confirmados y 0,08606 para fallecidos.



CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO QUE MIDE:

Nº	DIMENSIONES / ítems	Pertinencia ¹		Relevancia ²		Claridad ³		Sugerencias
		Si	No	Si	No	Si	No	
1	INDICADOR: Tasa de mortalidad (por cada 10000 habitantes) $TM = \frac{NTM}{NH} \times 10000$	X		X		X		
2	INDICADOR: Tasa de letalidad $TL = \frac{TF}{TC} \times 100$	X		X		X		

Observaciones (precisar si hay suficiencia): _____

Opinión de aplicabilidad: **Aplicable [X]** **Aplicable después de corregir []** **No aplicable []**

Apellidos y nombres del juez validador. Mg. William Rodriguez **CC: 1098685961 Colombia**

Especialidad del validador: Mg. Inteligencia

¹Pertinencia: El ítem corresponde al concepto teórico formulado.

²Relevancia: El ítem es apropiado para representar al componente o dimensión específica del constructo

³Claridad: Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión

29 de noviembre del 2020



TABLA DE EVALUACIÓN DE EXPERTOS METODOLOGÍA DE DESARROLLO

Apellidos y Nombres del Experto:	William Steve Rodríguez Villamizar
Título y/o Grado Académico:	Mg. Inteligencia Artificial e Ingeniería electrónica
Doctor () Magister (X) Ingeniero () Licenciado () Otro ()	
Fecha:	29/11/2020

TESIS: Redes neuronales artificiales para la predicción de casos por Covid-19 usando datos abiertos en el Perú, 2020

Autores: Catunta Esquivel Saulo - Roncal Hidalgo Dalyuska Mavel

MUY MAL (1) MALO (2) REGULAR (3) BUENO (4) EXCELENTE (5)

Mediante la tabla de evaluación de expertos usted tiene la facultad de evaluar la metodología de desarrollo de software involucradas mediante una serie de preguntas con puntuaciones especificadas al final de la tabla. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de las preguntas.

ITEM	PREGUNTAS	METODOLOGÍA		
		SCRUM	XP	KANBAN
1	¿Qué metodología brinda un mejor modelo de conocimiento para el trabajo de investigación?	5	3	5
2	¿Qué metodología propone un ciclo de vida en donde se indican las fases, las actividades y los productos más relevantes en el trabajo de investigación?	5	2	3
3	¿Qué metodología está enfocado a proyectos y es más fácil de entender y más autoorganizado del equipo?	5	3	3
4	¿Qué metodología define claramente las reglas que se utilizaran en el sistema del trabajo de investigación?	4	3	4
5	¿Qué metodología tiene una estructura más jerárquica?	5	3	4
6	¿Qué metodología es más flexible?	4	3	3
7	¿Qué metodología cuenta con un énfasis una documentación de los procesos para el desarrollo del proyecto?	4	2	3
PUNTUACIÓN		32	20	25

SUGERENCIAS

Al definir los ítems de proyecto, tener en cuenta las variantes no controladas que pueden generar atrasos de los mismos, así si llegaren a aparecer estas variables de incertidumbre, se podría abordar y solucionarlo sin repercutir en atrasos del proyecto.

FIRMA DEL EXPERTO



TABLA DE VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS: Tasa de mortalidad
I. DATOS GENERALES

Apellidos y Nombres del Experto: **William Steve Rodriguez Villamizar**
 Título y/o Grado Académico: **Mg. Inteligencia Artificial e Ingeniería electrónica**

Doctor () Magister (**X**) Ingeniero () Licenciado () Otro ()

Universidad que labora: **Universidad Internacional de Valencia**
 Fecha: **29/11/2020**

TESIS: Redes neuronales artificiales para la predicción de casos por Covid-19 usando datos abiertos en el Perú, 2020

Autores: Catunta Esquivel Saulo - Roncal Hidalgo Dalyuska Mavel

Deficiente (0-20%) Regular(21-50%) Bueno(51-70%) Muy Bueno(71-80%) Excelente(81-100%)

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucradas mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100%. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

II. ASPECTOS DE VALIDACIÓN

INDICADOR	CRITERIO	VALORACIÓN				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.				80 %	
OBJETIVIDAD	Esta expresado en conducta observable.				80 %	
ACTUALIDAD	es adecuado el avance, la ciencia y tecnología.				80 %	
ORGANIZACIÓN	Existe una organización lógica.				80 %	
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.				80 %	
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.				80 %	
CONSISTENCIA	Está basado en aspectos teóricos y científicos.				80 %	
COHERENCIA	En los datos respecto al indicador.				80 %	
METODOLOGÍA	Responde al propósito de investigación.				80 %	
PERTENENCIA	El instrumento es adecuado al tipo de investigación.				80 %	
TOTAL					80 %	

III. PROMEDIO DE VALIDACIÓN

80%

IV. OPCIÓN DE APLICABILIDAD

- (**X**) El instrumento puede ser aplicado, tal como está elaborado
 El instrumento debe ser mejorado antes de ser
 () aplicado

FIRMA DEL EXPERTO



TABLA DE VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS: Tasa de letalidad

I. DATOS GENERALES

Apellidos y Nombres del Experto: William Steve Rodríguez Villamizar
 Título y/o Grado Académico: Mg. Inteligencia Artificial e Ingeniería electrónica

Doctor () Magister (X) Ingeniero () Licenciado () Otro ().....

Universidad que labora: Universidad Internacional de Valencia
 Fecha: 29/11/2020

TESIS: Redes neuronales artificiales para la predicción de casos por Covid-19 usando datos abiertos en el Perú, 2020

Autores: Catunta Esquivel Saulo - Roncal Hidalgo Dalyuska Mavel

Deficiente (0-20%) Regular(21-50%) Bueno(51-70%) Muy Bueno(71-80%) Excelente(81-100%)

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucradas mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100%. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

II. ASPECTOS DE VALIDACIÓN

INDICADOR	CRITERIO	VALORACIÓN				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.				80 %	
OBJETIVIDAD	Esta expresado en conducta observable.				80 %	
ACTUALIDAD	es adecuado el avance, la ciencia y tecnología.				80 %	
ORGANIZACIÓN	Existe una organización lógica.				80 %	
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.				80 %	
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.				80 %	
CONSISTENCIA	Está basado en aspectos teóricos y científicos.				80 %	
COHERENCIA	En los datos respecto al indicador.				80 %	
METODOLOGÍA	Responde al propósito de investigación.				80 %	
PERTENENCIA	El instrumento es adecuado al tipo de investigación.				80 %	
TOTAL					80 %	


III. PROMEDIO DE VALIDACIÓN

80 %

IV. OPCIÓN DE APLICABILIDAD

- (X) El instrumento puede ser aplicado, tal como está elaborado
 El instrumento debe ser mejorado antes de ser
 () aplicado

FIRMA DEL EXPERTO



CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO QUE MIDE:

N°	DIMENSIONES / ítems	Pertinencia ¹		Relevancia ²		Claridad ³		Sugerencias
		Si	No	Si	No	Si	No	
1	INDICADOR: Tasa de mortalidad por 10000 $TM = \frac{NTM}{NH} \times 10000$	X		X		X		
2	INDICADOR: Tasa de letalidad $TL = \frac{TF}{TC} \times 100$	X		X		X		

Observaciones (precisar si hay suficiencia): Si hay suficiencia

Opinión de aplicabilidad: **Aplicable [X]** **Aplicable después de corregir []** **No aplicable []**

Apellidos y nombres del juez validador: Mg. Even Deyser Pérez Rojas

DNI: 43776841

Especialidad del validador: Magister en Gestión de Tecnologías de la Información

¹**Pertinencia:** El ítem corresponde al concepto teórico formulado.

²**Relevancia:** El ítem es apropiado para representar al componente o dimensión específica del constructo

³**Claridad:** Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión

12 de diciembre del 2020

Firma del Experto Informante.

TABLA DE EVALUACIÓN DE EXPERTOS METODOLOGÍA DE DESARROLLO

Apellidos y Nombres del Experto:	Mg. Even Deyser Pérez Rojas
Título y/o Grado Académico:	Magister en Gestión de tecnologías de información
Doctor () Magister () Ingeniero () Licenciado () Otro () Fecha:	11/12/2020

TESIS: Redes neuronales artificiales para la predicción de casos por Covid-19 usando datos abiertos en el Perú, 2020

Autores: Catunta Esquivel Saulo - Roncal Hidalgo Dalyuska Mavel

MUY MAL (1) MALO (2) REGULAR (3) BUENO (4) EXCELENTE (5)

Mediante la tabla de evaluación de expertos usted tiene la facultad de evaluar la metodología de desarrollo de software involucradas mediante una serie de preguntas con puntuaciones especificadas al final de la tabla. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de las preguntas.

ÍTEM	PREGUNTAS	METODOLOGÍA		
		SCRUM	XP	KANBAN
1	¿Qué metodología brinda un mejor modelo de conocimiento para el trabajo de investigación?	5	5	4
2	¿Qué metodología propone un ciclo de vida en donde se indican las fases, las actividades y los productos más relevantes en el trabajo de investigación?	5	4	5
3	¿Qué metodología está enfocado a proyectos y es más fácil de entender y más autoorganizado del equipo?	5	5	4
4	¿Qué metodología define claramente las reglas que se utilizaran en el sistema experto del trabajo de investigación?	5	4	5
5	¿Qué metodología tiene una estructura más jerárquica?	5	4	4
6	¿Qué metodología es más flexible?	5	5	5
7	¿Qué metodología cuenta con un énfasis una documentación de los procesos para el desarrollo del proyecto?	5	4	4
PUNTUACIÓN		35	31	31

SUGERENCIAS

FIRMA DEL EXPERTO





TABLA DE VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS: Tasa de mortalidad

I. DATOS GENERALES

Apellidos y Nombres del Experto:	Mg. Even Deyser Pérez Rojas
Título y/o Grado Académico:	Magister en Gestión de tecnologías de información

Doctor () Magister (X) Ingeniero () Licenciado () Otro ()

Universidad que labora:	Universidad César Vallejo
Fecha:	11/12/2020

TESIS: Redes neuronales artificiales para la predicción de casos por Covid-19 usando datos abiertos en el Perú, 2020

Autores: Catunta Esquivel Saulo - Roncal Hidalgo Dalyuska Mavel

Deficiente (0-20%) Regular(21-50%) Bueno(51-70%) Muy Bueno(71-80%) Excelente(81-100%)

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucradas mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100%. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

II. ASPECTOS DE VALIDACIÓN

INDICADOR	CRITERIO	VALORACIÓN				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.					90%
OBJETIVIDAD	Esta expresado en conducta observable.					90%
ACTUALIDAD	es adecuado el avance, la ciencia y tecnología.					90%
ORGANIZACIÓN	Existe una organización lógica.					90%
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.					90%
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.					90%
CONSISTENCIA	Está basado en aspectos teóricos y científicos.					90%
COHERENCIA	En los datos respecto al indicador.					90%
METODOLOGÍA	Responde al propósito de investigación.					90%
PERTENENCIA	El instrumento es adecuado al tipo de investigación.					90%
TOTAL						90%

III. PROMEDIO DE VALIDACIÓN

90%

IV. OPCIÓN DE APLICABILIDAD

- () El instrumento puede ser aplicado, tal como está elaborado
 El instrumento debe ser mejorado antes de ser
 () aplicado

FIRMA DEL EXPERTO

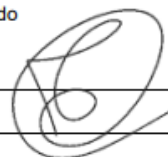


TABLA DE VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS: Tasa de letalidad
I. DATOS GENERALES

Apellidos y Nombres del Experto: **Mg. Even Deyser Pérez Rojas**
 Título y/o Grado Académico: **Magister en Gestión de tecnologías de información**

Doctor () Magister (**X**) Ingeniero () Licenciado () Otro ()

Universidad que labora: **Universidad César Vallejo**
 Fecha: **11/12/2020**

TESIS: Redes neuronales artificiales para la predicción de casos por Covid-19 usando datos abiertos en el Perú, 2020

Autores: Catunta Esquivel Saulo - Roncal Hidalgo Dalyuska Mavel

Deficiente (0-20%) Regular(21-50%) Bueno(51-70%) Muy Bueno(71-80%) Excelente(81-100%)

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucradas mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100%. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

II. ASPECTOS DE VALIDACIÓN

INDICADOR	CRITERIO	VALORACIÓN				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.					90%
OBJETIVIDAD	Esta expresado en conducta observable.					90%
ACTUALIDAD	es adecuado el avance, la ciencia y tecnología.					90%
ORGANIZACIÓN	Existe una organización lógica.					90%
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.					90%
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.					90%
CONSISTENCIA	Está basado en aspectos teóricos y científicos.					90%
COHERENCIA	En los datos respecto al indicador.					90%
METODOLOGÍA	Responde al propósito de investigación.					90%
PERTENENCIA	El instrumento es adecuado al tipo de investigación.					90%
TOTAL						

III. PROMEDIO DE VALIDACIÓN

90%

IV. OPCIÓN DE APLICABILIDAD

- () El instrumento puede ser aplicado, tal como está elaborado
 El instrumento debe ser mejorado antes de ser
 () aplicado

FIRMA DEL EXPERTO



CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO QUE MIDE:

N°	DIMENSIONES / Items	Pertinencia ¹		Relevancia ²		Claridad ³		Sugerencias
		Si	No	Si	No	Si	No	
1	INDICADOR: Tasa de mortalidad por 100000							
	$TM = \frac{NTM}{NH} \times 100000$	X		X		X		
2	INDICADOR: Tasa de letalidad							
	$TL = \frac{TF}{TC} \times 100$	X		X		X		

Observaciones (precisar si hay suficiencia): _____

Opinión de aplicabilidad: **Aplicable [X]** **Aplicable después de corregir []** **No aplicable []**

Apellidos y nombres del juez validador: Mag.Ing. Liendo Arevalo Milner David

DNI: 00792777

Especialidad del validador: Magister

29 de noviembre del 2020

¹Pertinencia: El ítem corresponde al concepto teórico formulado.

²Relevancia: El ítem es apropiado para representar al componente o dimensión específica del constructo

³Claridad: Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión



Firma del Experto Informante.

TABLA DE EVALUACIÓN DE EXPERTOS METODOLOGÍA DE DESARROLLO

Apellidos y Nombres del Experto:	Mag.Ing. Liendo Arevalo Milner David
Título y/o Grado Académico:	Doctora en educación
Doctor () Magister (x) Ingeniero () Licenciado () Otro ()	
Fecha:	29/11/2020

TESIS: Redes neuronales artificiales para la predicción de casos por Covid-19 usando datos abiertos en el Perú, 2020

Autores: Catunta Esquivel Saulo - Roncal Hidalgo Dalyuska Mavel

MUY MAL (1) MALO (2) REGULAR (3) BUENO (4) EXCELENTE (5)

Mediante la tabla de evaluación de expertos usted tiene la facultad de evaluar la metodología de desarrollo de software involucradas mediante una serie de preguntas con puntuaciones especificadas al final de la tabla. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de las preguntas.

ÍTEM	PREGUNTAS	METODOLOGÍA		
		SCRUM	XP	KANBAN
1	¿Qué metodología brinda un mejor modelo de conocimiento para el trabajo de investigación?	X		
2	¿Qué metodología propone un ciclo de vida en donde se indican las fases, las actividades y los productos más relevantes en el trabajo de investigación?	X		
3	¿Qué metodología está enfocado a proyectos y es más fácil de entender y más autoorganizado del equipo?	X		
4	¿Qué metodología define claramente las reglas que se utilizaran en el sistema experto del trabajo de investigación?	X		
5	¿Qué metodología tiene una estructura más jerárquica?	X		
6	¿Qué metodología es más flexible?	X		
7	¿Qué metodología cuenta con un énfasis una documentación de los procesos para el desarrollo del proyecto?	X		
PUNTUACIÓN		35		

SUGERENCIAS

FIRMA DEL EXPERTO





UNIVERSIDAD CÉSAR VALLEJO

TABLA DE VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS: Tasa de mortalidad

I. DATOS GENERALES

Apellidos y Nombres del Experto: Mag.Ing. Liendo Arevalo Milner David
Título y/o Grado Académico: Magister

Doctor () Magister (X) Ingeniero () Licenciado () Otro ()

Universidad que labora: Universidad César Vallejo
Fecha: 29/11/2020

TESIS: Redes neuronales artificiales para la predicción de casos por Covid-19 usando datos abiertos en el Perú, 2020

Autores: Catunta Esquivel Saulo - Roncal Hidalgo Dalyuska Mavel

Deficiente (0-20%) Regular(21-50%) Bueno(51-70%) Muy Bueno(71-80%) Excelente(81-100%)

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucradas mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100%. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

II. ASPECTOS DE VALIDACIÓN

INDICADOR	CRITERIO	VALORACIÓN				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.				80	
OBJETIVIDAD	Esta expresado en conducta observable.					90
ACTUALIDAD	es adecuado el avance, la ciencia y tecnología.					90
ORGANIZACIÓN	Existe una organización lógica.					90
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.					90
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.					90
CONSISTENCIA	Está basado en aspectos teóricos y científicos.					90
COHERENCIA	En los datos respecto al indicador.					90
METODOLOGÍA	Responde al propósito de investigación.					90
PERTENENCIA	El instrumento es adecuado al tipo de investigación.					90
TOTAL						

III. PROMEDIO DE VALIDACIÓN

IV. OPCIÓN DE APLICABILIDAD

- (X) El instrumento puede ser aplicado, tal como está elaborado
El instrumento debe ser mejorado antes de ser
() aplicado

FIRMA DEL EXPERTO

**TABLA DE VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS: Tasa de letalidad****I. DATOS GENERALES**

Apellidos y Nombres del Experto:	Mag.Ing. Liendo Arevalo Milner David
Título y/o Grado Académico:	Magister

Doctor () Magister (**X**) Ingeniero () Licenciado () Otro ()

Universidad que labora:	Universidad César Vallejo
Fecha:	29/11/2020

TESIS: Redes neuronales artificiales para la predicción de casos por Covid-19 usando datos abiertos en el Perú, 2020

Autores: Catunta Esquivel Saulo - Roncal Hidalgo Dalyuska Mavel

Deficiente (0-20%) Regular(21-50%) Bueno(51-70%) Muy Bueno(71-80%) Excelente(81-100%)

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucradas mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100%. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

II. ASPECTOS DE VALIDACIÓN

INDICADOR	CRITERIO	VALORACIÓN				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.				80	
OBJETIVIDAD	Esta expresado en conducta observable.					90
ACTUALIDAD	es adecuado el avance, la ciencia y tecnología.					90
ORGANIZACIÓN	Existe una organización lógica.					90
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.					90
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.					90
CONSISTENCIA	Está basado en aspectos teóricos y científicos.					90
COHERENCIA	En los datos respecto al indicador.					90
METODOLOGÍA	Responde al propósito de investigación.					90
PERTENENCIA	El instrumento es adecuado al tipo de investigación.					90
TOTAL						

III. PROMEDIO DE VALIDACIÓN

--

IV. OPCIÓN DE APLICABILIDAD

- () El instrumento puede ser aplicado, tal como está elaborado
 El instrumento debe ser mejorado antes de ser
 () aplicado

FIRMA DEL EXPERTO



COMPARATIVO ENTRE METODOLOGÍAS

Si comparamos KDD y SEMMA podemos afirmar que son equivalentes:

Podemos decir que las cinco etapas del proceso SEMMA pueden ser vistas como una implementación práctica de las cinco etapas del proceso KDD, ya que está directamente vinculada al software SAS Enterprise Miner.

TABLA 47 COMPARACIÓN KDD - SEMMA

KDD	SEMMA
Muestra	Selección.
Explorar	Pre-procesamiento
Modificar	Transformación
Modelo	Data Mining
Evaluación	Interpretación / Evaluación

Si comparamos KDD y CRISP-DM podemos observar que la metodología CRISP-DM incorpora los pasos que deben preceder y seguir el proceso KDD:

TABLA 48 COMPARACIÓN KDD - CRISP-DM

KDD	CRISP-DM
La fase de Entendimiento de Negocios	Puede identificarse con el desarrollo de una comprensión del dominio de la aplicación, el conocimiento previo relevante y los objetivos del usuario final.
La fase de implementación	Puede identificarse con la consolidación incorporando este conocimiento en el sistema.
La fase de Entendimiento de Datos	Puede ser identificada como la combinación de Selección y Pre procesamiento.
La fase de preparación	Se puede identificar con Transformación.
La fase de modelado	Se puede identificar con Data Mining
La fase de Evaluación	Puede ser identificada con Interpretación/ Evaluación.

Fuente: Grández Márquez Miguel

Comparación CRISP –DM Y SEMMA

TABLA 49 COMPARACIÓN CRIP-DM - SEMMA

SEMA	CRIP-DM
Muestra y Exploración	Fase de Entendimiento de Datos
Modificar	Fase de preparación de datos
Modelo	Fase de modelado,
Evaluar paralelos	Fase de evaluación
Ambos modelos pretenden ser algo cíclicos en lugar de lineales.	

Fuente: Grández Márquez Miguel

Diferencias entre CRISP-DM y SEMMA. 22

SEMMA fue desarrollado para un paquete específico de software: SAS Enterprise Miner y, pone menos énfasis en las fases de planificación inicial cubiertas en las fases de CRISP-DM (Business Understanding y Data Understanding) y omite totalmente la fase de implementación.

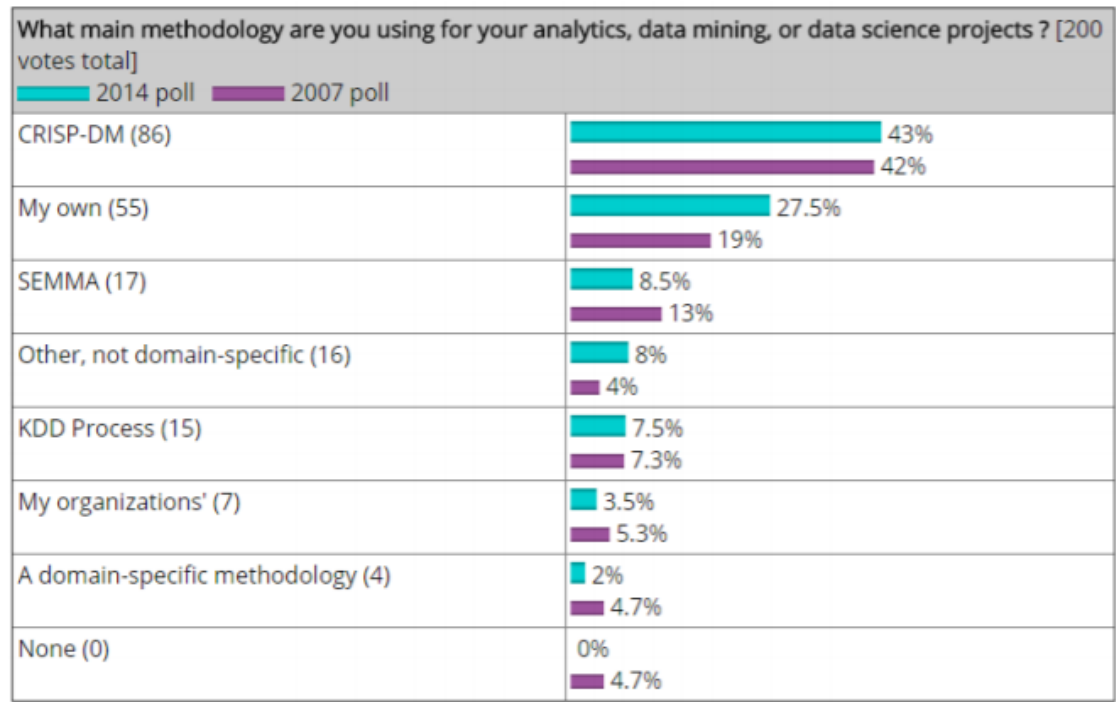
ELECCIÓN DE LA METODOLOGÍA

Para elegir la metodología a implementar realizamos un análisis sobre la base de características deseadas, luego de lo cual se elige implementar la metodología que satisface de forma completa los criterios de selección.

Deseamos una metodología que sea de amplio uso entre profesionales por lo cual recurrimos a un análisis de kdnuggets en el que se puede ver la preferencia de 200 usuarios en el uso de una metodología de minería de datos.

Vemos que en primer lugar de preferencia se encuentra la metodología CRISP –DM con un 43%, en segundo lugar SEMMA con un 8.5% y en tercer lugar KDD con 7.5%.

FIGURA 57 PREFERENCIA DE METODOLOGÍA



Fuente: Grández Márquez Miguel

Además de ser una metodología predominante en su uso, también deseamos otros aspectos.